Mixing it up: Inflation at risk*

Maximilian Schröder[1,2]

[1]European Central Bank

[2]BI Norwegian Business School

November 15, 2025

Abstract

Understanding how risk factors shape the economic outlook is essential for guiding policy decisions. This paper develops a flexible framework that analyses how economic indicators influence the entire forecast distribution of macroeconomic variables. The framework decomposes distributional risk forecasts into their underlying predictors and supports the construction of interpretable risk measures. Multiple modelling strategies, including density and quantile regression, are accommodated underscoring the versatility of the approach. The framework is illustrated using recent U.S. inflation data, showing that post-pandemic risk forecasts were driven by the business cycle, commodity prices, monetary policy, and inflation expectations. Simulation studies and empirical analyses demonstrate robustness across models and sample sizes, establishing the framework as a general tool for macroeconomic and financial risk assessment.

Keywords: Bayesian methods, density regression, quantile regression, macroeconomic risk, inflation, inflation risk

JEL Classification: C11, C22, C51, C52, C53, E6

*I would like to thank two anonymous referees, Sylvia Kaufmann, Sylvia Frühwirth-Schnatter, Hilde Bjørnland, Silvia Miranda-Agrippino, Dimitris Korobilis, Leif A. Thorsrud, Mirco Rubin, Jim Griffin, Fabio Canova, Michele Lenza, Franceso Ravazzolo, Francesco Furlanetto, Domenico Giannone, Giorgio Primiceri, Herman van Dijk, Ørjan Robstad, Lennart Brandt, Frank Schorfheide, and participants at the following conferences for useful comments and discussions: 12th European Seminar on Bayesian Econometrics in Glasgow, 3rd Sailing the Macro Workshop in Siracusa, 3rd Dolomiti Macro Meeting in San Candido, Junior Workshop in Econometrics and Applied Economics in Rome, 31st SNDE Symposium 2024 in Padova, Workshop in Empirical Macroeconomics in Innsbruck. I would also like to thank seminar participants at the following institutions: BI Norwegian Business School; Norges Bank. This paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB. This paper was part of the research activities at the Centre for Applied Macroeconomics and Commodity Prices (CAMP) at the BI Norwegian Business School.

1 Introduction & Motivation

Monetary policy is inherently forward-looking and relies on forecasts that are subject to substantial uncertainty, making risk management a core element of modern central banking (Greenspan, 2004; Kilian and Manganelli, 2008). Episodes such as the Great Recession, the COVID-19 pandemic, and the recent inflation surge illustrate how difficult this task becomes when major risk dynamics emerge. Global supply chain disruptions, energy market shocks, and the post-lockdown rebound in demand, amplified by the lagged effects of expansionary fiscal and monetary policy, pushed prices upward worldwide and left policymakers uncertain about whether price pressures would prove temporary or persistent and whether upward or downward pressures would prevail.

Comparable difficulties extend to other macroeconomic indicators, including output growth and employment and underscore a broader challenge: while substantial progress has been made in quantifying forecast uncertainty, understanding its sources remains difficult. The same level of macroeconomic risk can arise from very different underlying dynamics, each calling for a distinct policy response. Turning such abstract measures of uncertainty into actionable insight therefore requires decomposing the risk outlook into its underlying contributions.

This paper develops a unified framework that directly addresses this challenge. Operating on the full predictive density without imposing distributional restrictions, it can be readily embedded in existing forecasting setups. The proposed algorithms attribute movements in the predictive distribution, and in the risk measures derived from it, to economic variables, providing a coherent and interpretable view of the macroeconomic risk outlook through the lens of at-risk forecasting models.

As a starting point, the paper builds on two complementary approaches for modelling predictive distributions. The first is quantile regression (QR) (Koenker and Bassett, 1978), the workhorse of the "macroeconomics-at-risk" literature (e.g. Adrian et al., 2019; Lopez-Salido and Loria, 2024; Mitchell et al., 2024; Tagliabracci, 2020), which models conditional quantiles which are then combined into densities without parametric restrictions. The second is density regression (DR), following Rigon and Durante (2021), which employs flexible mixture models that capture skewness, heavy tails, and multimodality (Villani et al., 2009; Frühwirth-Schnatter, 2006) and model the forecast density directly. Together, these approaches offer a data-driven representation of forecast uncertainty, providing the foundation for the decomposition framework developed later in the paper. The ability of both models to approximate complex distributional shapes is illustrated through simulation studies and a forecasting exercise

on U.S. inflation, demonstrating that they perform robustly across a wide range of conditions, particularly during periods of elevated uncertainty.

Building on these models, the paper introduces decomposition algorithms. Grounded in Shapley value theory (Shapley, 1953) and drawing on advances in explainable machine learning (Strumbelj and Kononenko, 2010), the methods attribute shifts in local probability mass to individual predictors, translating complex statistical forecasts into interpretable economic contributions. The algorithms are fully model-independent and integrate seamlessly with existing forecasting frameworks, providing a consistent and transparent way to analyse macroeconomic risk.

As an additional extension and validation, the framework is applied to the construction and analysis of preference-consistent risk measures. Following Kilian and Manganelli (2007, 2008), these measures summarize the predictive density in a way that reflects preferences towards downside or upside risks, linking empirical forecasts to established decision-theoretic concepts in monetary policy. Because they are derived directly from the estimated distributions, they remain consistent with the underlying model and capture asymmetries and tail behaviour. Moreover, since the proposed decomposition methods extend to any non-linear transformation of the density, the risk measures can be decomposed within the same unified framework, providing a link between statistical representations of uncertainty, summary measures derived from them, and their economic, albeit non-causal, interpretation.

The performance and interpretability of the proposed method are demonstrated in a series of simulation exercises. These experiments show that the Shapley-based algorithms provide internally consistent and robust decompositions, and that the results remain stable across model types and sample sizes. Together, these simulations confirm that the framework performs reliably across a wide range of empirical environments, laying the groundwork for its application to diverse macroeconomic settings.

To illustrate this, the framework is applied to analyse U.S. inflation risks throughout the post-pandemic period. The decomposition results are consistent across both density forecasting models and reveal four distinct phases: an initial phase coinciding with the global rebound in demand following the pandemic; a second phase characterized by elevated commodity prices, loose financial conditions, and monetary policy - patterns consistent with global demand dynamics, supply-side shocks, and the lagged effects of earlier policy easing; a third phase associated with persistent inflation and emerging downward pressures from monetary tightening; and a final phase in which elevated inflation expectations maintained upward pressure as other influences receded.

These results align closely with the structural and narrative evidence reported by

Abbai et al. (2022), Ball et al. (2022), Bai et al. (2024), Blanchard and Bernanke (2023), Cascaldi-Garcia et al. (2024), Giannone and Primiceri (2024), and Ha et al. (2021), indicating that the framework captures key macroeconomic mechanisms as they evolved in real time and offers an initial empirical validation of the proposed decomposition algorithms. Given that the analysis is reduced-form rather than causal, a complementary exercise relates the contributions of energy prices and global business cycle dynamics to the structural shocks identified by Baumeister and Hamilton (2019), revealing relatively strong correlations. Taken together, these results show that the framework yields a realistic and interpretable representation of macroeconomic risk by linking movements in the forecast distribution directly to their underlying predictors.

The remainder of the paper is structured as follows. Section 2 outlines the modelling setup, detailing the quantile and density regression models and their estimation. Section 3 introduces the decomposition algorithms and discusses their implementation and validation through simulation studies. Section 4 presents the empirical application, and section 5 extends the analysis to risk measures, illustrating the evolution and composition of U.S. inflation risks. Section 6 concludes.

2 Models and Estimation

The distributions of macroeconomic and financial variables often display pronounced non-Gaussian features, such as skewness, fat tails, and multimodality. While these patterns are readily visible in the unconditional data[1], the macroeconomic-uncertainty literature shows that similar characteristics apply to the conditional forecast distributions of macroeconomic aggregates. For GDP growth, the conditional distribution has been observed to skew left during recessions, co-moving with financial conditions and macroeconomic volatility (Adrian et al., 2019; Bekaert and Popov, 2019; Monache et al., 2024). For inflation, asymmetries and skewness are likewise observed by Banerjee et al. (2024), Korobilis et al. (2021), Korobilis (2017), Lopez-Salido and Loria (2024), and Tagliabracci (2020), with similar evidence reported for other macroeconomic and financial time series (Korobilis and Schröder, 2025a). These findings motivate quantile and density regression models, introduced in this section, as flexible frameworks for characterising the full conditional distribution of macroeconomic outcomes.

[1]Descriptive evidence for the unconditional distribution of U.S. PCE excluding energy and food inflation is provided in Figure 2 of the online appendix.

2.1 Quantile Regression

In the macroeconomic uncertainty literature, quantile regression has become a standard tool for forecasting conditional percentiles of economic variables. The q-th conditional quantile of a variable y_{t+h} at forecast horizon h is related to a vector $\boldsymbol{x}_t$, collecting the intercept and predictors, through

$$Q_q(y_{t+h}|\boldsymbol{x}_t) = \boldsymbol{\beta}_q \boldsymbol{x}_t + u_{t,q}, \tag{1}$$

for quantile level $q \in [0, 1]$. In a Bayesian setting, quantile estimates can be obtained by adopting the asymmetric Laplace (AL) likelihood (Yu and Moyeed, 2001),

$$AL_q(u_t; 0, \sigma_q) = \frac{q(q-1)}{\sigma_q} exp\Big\{-\frac{1}{\sigma_q}\rho_q(u_{t,q})\Big\}, \tag{2}$$

where σ_q is a scale parameter and $\rho_q(u) = u(q - \mathbb{I}u \leq 0)$ denotes the check loss function familiar from frequentist QR (Koenker, 2005). For practical inference, the Gaussian scale-mixture representation of the AL distribution (Kozumi and Kobayashi, 2011) is typically exploited, yielding the conditionally Gaussian form

$$u_{t,q} = \frac{1-2q}{1-q} z_{t,q} + \sqrt{\frac{2\sigma_q z_{t,q}}{q(1-q)}} v_{t,q}, \tag{3}$$

with $z_{t,q} \sim \text{Exp}(\sigma_q)$ and $v_{t,q} \sim N(0, 1)$. This representation allows standard priors and samplers to be used efficiently. To approximate the entire forecast distribution, a grid of quantile regressions is often estimated and then matched to a flexible parametric density, such as the skewed-t distribution (Adrian et al., 2019; Lopez-Salido and Loria, 2024). While widely used, this approach can over-smooth the estimated densities, potentially obscuring features such as multimodality. Mitchell et al. (2024) address this concern by reconstructing the density directly from a fine grid of quantile forecasts by interpolating across adjacent quantiles. For $q_{j-1} < q \leq q_j$, values are sampled using

$$\hat{\boldsymbol{y}}_{t+h,j} = \beta_{q_{j-1}} \boldsymbol{x}_t \mathbf{1}_{(q_j - q_{j-1})N} + (\beta_{q_j} \boldsymbol{x}_t - \beta_{q_{j-1}} \boldsymbol{x}_t) \boldsymbol{U}_j, \tag{4}$$

where $\mathbf{1}_{(q_j - q_{j-1})N}$ and $\boldsymbol{U}_j$ are $(q_j - q_{j-1})N$ vectors of ones and standard uniform draws, respectively. The full density is recovered by aggregating all N draws across all intervals.

2.2 Density Regression

As a complementary approach, this paper adapts the density regression model of Rigon and Durante (2021) to directly estimate conditional forecast densities of macroeconomic variables. In this framework, the conditional density at horizon h is expressed as

$$f_x(y_{t+h}) = \int K_{\boldsymbol{x}}(y_{t+h};\boldsymbol{\theta})p_{\boldsymbol{x}}(d\boldsymbol{\theta}) = \sum_{c=1}^{\infty} w_c(\boldsymbol{x}_t)K_{\boldsymbol{x}}(y_{t+h};\boldsymbol{\theta}_c), \tag{5}$$

where $K_{\boldsymbol{x}}(y_{t+h};\boldsymbol{\theta})$ denotes a predictor dependent parametric kernel with parameters $\boldsymbol{\theta}$, p denotes a probability measure, and $w_c(\boldsymbol{x}_t) = \nu_c(\boldsymbol{x}_t)\prod_{l=1}^{c-1}(1-\nu_l(\boldsymbol{x}_t))$ are predictor dependent mixture probabilities with $w_1(\boldsymbol{x}_t) = \nu_1(\boldsymbol{x}_t)$ defined by stick-breaking weights $\nu_c(\boldsymbol{x}_t)$ for every $c \geq 1$. In this application, a Gaussian kernel is used, $K_{\boldsymbol{x}}(y_{t+h};\boldsymbol{\theta}_c) = \sqrt{\tau_c}\phi\left(\sqrt{\tau_c}(y_{t+h} - \boldsymbol{x}_t\boldsymbol{\beta}_c)\right)$, where $\tau_c = \sigma_c^{-2}$ and $\boldsymbol{\beta}_c$ are component-specific precision parameters and regression coefficients, respectively. Each component thus shares the same predictive structure but with potentially different parameters. This way, the forecast distribution is linked to economic conditions through the non-linear interaction of mixture components and weights. Intuitively, this resembles a regime-switching model with infinitely many regimes, where both the regimes and transition probabilities depend directly on data.[2]

This stands in contrast to Labonne (2025) and Monache et al. (2024) who model the distribution of GDP growth using a flexible skewed-t distribution with data-dependent moments that accommodates skewness and heavy tails. While effective, this parametric family is unimodal and may smooth over potential multimodality. Density regression instead, can capture multimodality and could be viewed as a more structural alternative to the quantile-based density estimation method in Mitchell et al. (2024).

To achieve this flexibility in a computationally tractable way, Rigon and Durante (2021) introduce a logistic stick breaking prior that links the stick-breaking weights, $\nu_c(\boldsymbol{x}_t)$, directly to the variables via the logistic function. Writing $\eta_c(\boldsymbol{x}_t) = logit(\nu_c(\boldsymbol{x}_t)) = log(\nu_c(\boldsymbol{x}_t)/(1-\nu_c(\boldsymbol{x}_t)))$, the mixture probabilities $w_c(\boldsymbol{x}_t)$ are given by

$$w_c(\boldsymbol{x}_t) = \frac{exp\left\{\eta_c(\boldsymbol{x}_t)\right\}}{1+exp\left\{\eta_c(\boldsymbol{x}_t)\right\}}\prod_{l=1}^{c-1}\left[\frac{1}{1+exp\left\{\eta_l(\boldsymbol{x}_t)\right\}}\right]. \tag{6}$$

[2]The model's validity in time-series settings relies on the standard assumption of conditional independence given the lagged predictors x_t. As in time-series quantile regression, temporal dependence is captured through the lag structure and, in this case, also through the covariate-dependent mixing weights $w_c(x_t)$. For asymptotic inference, the underlying process is assumed to be stationary and weakly dependent (e.g., ρ-mixing).

Here, $\eta_c(\boldsymbol{x}_t) = \boldsymbol{\psi}_c \boldsymbol{x}_t$, is the log-odds of the probability of observation t being allocated to component c, given that it has not been allocated to any of the previous components. This breaks up the multinomial choice problem into a sequence of binary logistic regressions.

To simplify the problem further, Bayesian estimation can leverage the Pólya-Gamma augmentation in Polson et al. (2013). Conditioning on $\omega_{t,c} \sim PG(1, \boldsymbol{\psi}_c \boldsymbol{x}_t)$ yields a conditionally Gaussian form, reducing each logistic regression to a linear regression of the transformed assignment decision indicators $(z_{t,c} - 0.5)/\omega_{t,c}$ onto the predictors $\boldsymbol{x}_t$. This approach is tuning-free, computationally fast, and typically more efficient than Metropolis-Hastings alternatives.

Unlike quantile regression, which requires post-processing of posterior estimates to reconstruct the density, the density regression framework yields the forecast distribution directly in closed form:

$$f_x(y)_{t+h|t} = \sum_{c=1}^{\infty} \left[\nu_c(\boldsymbol{x}_t) \prod_{l=1}^{c-1} \{1 - \nu_l(\boldsymbol{x}_t)\} \right] \cdot \sqrt{\tau_c} \phi \left(\sqrt{\tau_c} (y - \boldsymbol{x}_t \beta_c) \right). \tag{7}$$

Both frameworks will be employed in the following sections to develop decomposition algorithms for density forecasts and risk measures.

2.3 Shrinkage and Priors

A key advantage of Bayesian inference is the use of flexible prior distributions that provide automatic regularization of the likelihood and enforce parsimony. This is particularly valuable when covariate sets are large and the available data limited.

In quantile regression, for instance, shrinkage priors help improve estimates for extreme quantiles, where inference relies on only a few observations (Korobilis and Schröder, 2025a). In density regression, a similar need arises when mixing probabilities and components depend on many predictors or when individual components contain only a small number of observations. In such cases, penalizing complexity improves parameter estimates and predictive accuracy (Ding and Karabatsos, 2021; Malsiner-Walli et al., 2016).

To address this, this paper imposes the horseshoe prior of Carvalho et al. (2010):

$$\{\boldsymbol{\beta}_q, \boldsymbol{\psi}_c\} : \mathcal{B} | \xi, \bar{\boldsymbol{\tau}} \sim \prod_{j=1}^{k(\mathcal{B})} N(0, \xi \bar{\tau}_j), \text{ for } \xi, \bar{\tau}_j \sim C^+(0, 1) \tag{8}$$

where C^+ denotes the half-Cauchy distribution and $k(\cdot)$ denotes the vector length. The prior is imposed on the regression parameters $\boldsymbol{\beta}_q$ in the quantile regression and on the log-odds parameters $\boldsymbol{\psi}_c$ in the density regression. Under this prior, the posterior remains largely unrestricted in small parameter spaces, but increasing shrinkage is automatically imposed as the parameter space grows relative to the sample size, making it the default choice in many applications (Korobilis and Shimizu, 2022). Due to the conditionally normal form of the quantile regression and the individual logistic regressions governing the mixture weights, the prior can be imposed without adjustments and sampling proceeds as in standard regression models. Its application in density regression is further supported by Bhattacharyya et al. (2022) and Wei and Ghosal (2020), who show that the favourable properties of the horseshoe prior extend to Bayesian logistic regression.

Additional shrinkage in the density regression framework arises from the logistic stick-breaking prior itself. Because the sequence of binomial choice problems terminates once an observation is allocated to a mixture component, later components may never be reached and thus receive zero probability. This mechanism limits the effective number of mixture components, implicitly penalizes small clusters, and reduces the risk of overfitting.[3]

The remaining priors are given by $\sigma_q \sim IG(r_{q,0}, s_{q,0})$ and $z_{q,t} \sim \text{Exp}(1)$ for the quantile regressions and by $\tau_c \sim IG(a_\tau, b_\tau)$, $\beta_c \sim N(0,1)$ for the kernel-specific precision parameters and regression coefficients in the density regression, respectively.

2.4 Estimation Algorithm

In this paper, the models are estimated using Variational Bayes (VB). Unlike MCMC, which provides an asymptotically exact representation of the posterior through sampling, VB minimizes the Kullback-Leibler divergence between the true posterior and a simpler distribution, typically factorized into independent variational families. This yields substantial computational gains, at the cost of only approximating the true posterior. VB is thus well suited for large-scale forecasting, real-time monitoring, and other high-dimensional tasks (for an overview, see Blei et al., 2017).

In the density regression framework, VB offers an additional advantage through its robustness to label switching. While MCMC posterior draws may require reordering if interest lies in component-specific parameters rather than the forecast density itself, VB converges to a single solution, avoiding the need for post-processing.

[3]In practice, the algorithm operates on a truncated version of the probability measure that serves as an approximation to the infinite mixture model. For a truncation point C, only the first $C-1$ components are modelled and $\nu_C = 1$. Rigon and Durante (2021) show that this approximation is accurate, as long as the truncation point is set sufficiently large. In all empirical exercises, $C = 5$ is assumed.

For quantile regression, the VB algorithm follows Lim et al. (2020) and cycles through:

1. Update the quantile regression coefficients β_q and corresponding horseshoe prior parameters, using the conjugate parametrization of Makalic and Schmidt (2016).[4]
2. Update the auxiliary parameters $z_{q,t}$ and the scale parameters σ_q.

Correspondingly, the updating scheme for the density regression model closely follows the VB algorithm in Rigon and Durante (2021), cycling through:

1. Update the binary mixture assignment indicators, $z_{t,c}$, and Pólya-Gamma distributed auxiliary variables, $\omega_{t,c}$.
2. Exploiting the Pólya-Gamma augmentation of Polson et al. (2013), update the parameters of the logistic regressions, $\boldsymbol{\psi}_c$, that govern component allocation, via Bayesian linear regression of the shifted assignment indicators, $(z_{t,c} - 0.5)$, on $\boldsymbol{x}_t$.
3. Update the corresponding parameters of the horseshoe prior based on the hierarchical parametrization of Makalic and Schmidt (2016).
4. Update the kernel specific regression and precision parameters, $\boldsymbol{\beta}_c$ and τ_c, exploiting standard Bayesian linear regression results.

Further estimation details are provided in Online Appendix 2.

3 Decomposing the Forecast Density

The preceding sections presented different econometric frameworks for constructing forecast densities that can accommodate complex distributional features. While these models provide flexible estimates of conditional distributions, they do not in themselves reveal how the shape of the forecast density arises from the underlying data. Likewise, their parameters provide only a partial view, as they do not explain how predictors collectively generate asymmetries, skewness, tail risks, or multimodality. The decomposition framework developed below addresses this gap by attributing changes in the density's shape to individual predictors in an additive and comprehensive way.

Developing such a framework is challenging because forecast densities are inherently complex, non-linear objects. In the case of point forecasts obtained from a linear regression, each coefficient directly measures the contribution of a variable to the conditional mean, and these contributions remain independent of changes in other

[4]This parametrization leads to conjugate posteriors, enabling fast and simple updates.

variables. For densities, however, this separability no longer holds. Even in symmetric and unimodal cases, the slope of the conditional density varies across its support, so that a change in one variables alters not only its own contribution to local probability mass but also the contributions of other variables. With more flexible densities, whether constructed from mixtures or by interpolating across neighbouring quantiles, additional non-linearities arise. In case of density regression for example, an increase in variable x_j may alter the contribution of variable x_i solely through changes in the mixture weights and, consequently, the overall composition of the density.

In motivating the framework, the first step is to simplify the problem - without loss of generality - by discretizing the density into arbitrarily fine bins of probability mass. The problem now becomes measuring how changes in the data reallocate probability mass across bins as the density evolves.

As starting point one might consider to compute partial gradients as the data moves from some point $\boldsymbol{x}$ to a point $\boldsymbol{x}'$. While intuitive, gradients are inherently local. They measure changes in probability mass for infinitesimal perturbations, but when used to approximate the total shift between $\boldsymbol{x}'$ and $\boldsymbol{x}$, they effectively extrapolate a straight line. Unless the relationship is globally linear, this deviation from the true trajectory leads to aggregation errors in which contributions do not sum consistently across variables or bins.

A natural extension is to accumulate gradients along the path from $\boldsymbol{x}$ to $\boldsymbol{x}'$, defined by $\gamma(\alpha) = \boldsymbol{x} + \alpha(\boldsymbol{x}' - \boldsymbol{x})$ for $\alpha \in [0, 1]$, rather than relying on a single local slope. This yields integrated gradients (IG), which, for sufficiently small intervals, ensure that contributions add up exactly to the total change in probability mass, thereby resolving the aggregation problem (Sundararajan et al., 2017). Nonetheless, IG restricts attention to a single straight-line path between $\boldsymbol{x}$ and $\boldsymbol{x}'$, even though infinitely many alternative paths exist and may in general produce different attributions.

Shapley values, dating back to Shapley (1953), extend the path-based intuition by averaging contributions across all possible paths of variable inclusions from the empty coalition, where no predictors are included, to the observed data point (Sundararajan et al., 2017). Instead of relying on a single trajectory, as in integrated gradients, every ordering of predictors is considered, and the marginal effect of each variable is recorded as it enters. Averaging these effects across all orderings ensures path independence. The idea originates in cooperative game theory, where the payoff of a game - here, the prediction of a model - is distributed fairly among the players, in this case the variables (Strumbelj and Kononenko, 2010). This yields a unique allocation of contributions that satisfies well-known fairness axioms, ensuring a principled decomposition of model outputs.[5]

[5]Integrated Gradients satisfy implementation invariance, sensitivity, linearity, and completeness, and

Extending these ideas to the present context, the same principle yields a unique, additive decomposition of changes in the forecast density across predictors.

Definition 3.1 (*Contributions to the Forecast Distribution*)

The contribution of variable x_j to the forecast distribution at time t for horizon $t+h$ in the interval $y^\star_{i-1}$ to $y^\star_i$, is

$$\varphi_j^{y_i^\star} = \int_{y^\star_{i-1}}^{y^\star_i} f_x(y)_{t+h|t} dy - \int_{y^\star_{i-1}}^{y^\star_i} f_x(y)_{t+h} dy - \sum_{l \in [1,\dots,n]: \neq j} \varphi_l^{y_i^\star}. \tag{9}$$

Then, x_j contributes to variation of the forecast distribution, if $\varphi_j^{y_i^\star} \neq 0$.

This definition implies that the sum of all contributions, $\sum_l^n \varphi_l^{y_i^\star}$, in probability bin $y^\star_{i-1}$ to $y^\star_i$, is precisely equal to the difference between the probability mass of the forecast distribution and the probability mass of the average across all historical forecast distributions in that bin. The historical average thereby plays the role of the empty coalition payoff in the Shapley construction and provides a natural normalization. Viewed through the Shapley fairness axioms, several properties follow: contributions add up exactly in each bin (efficiency), constants receive zero attribution (dummy), decompositions remain linear regardless of model complexity (linearity), group contributions can be obtained by summing individual ones (additivity), and predictors that enter symmetrically are treated equally (symmetry). Finally, with this definition, the individual contributions to the forecast distribution satisfy Lemma 3.1.

Lemma 3.1

Under definition 3.1 and given g grid points $y^\star$, such that $y^\star_{i-1} < y^\star_i$, the contributions of the individual predictors preserve integrability, i.e.

$$\sum_{i=2}^{g} \left[\sum_{j}^{n} \varphi_j^{y_i^\star} + \int_{y^\star_{i-1}}^{y^\star_i} f_x(y)_{t+h} dy \right] = \int_{-\infty}^{\infty} f_x(y)_{t+h|t} dy = 1.$$

The proof is given in Online Appendix 1.

They are hence directly comparable across forecast horizons, time periods, and probability bins.

are the unique symmetry-preserving path method along the straight-line path from baseline to input (Sundararajan et al., 2017). Shapley values extend these principles by averaging over all possible paths (predictor coalitions), which adds the game-theoretic axioms of symmetry and dummy (Shapley, 1953).

3.1 Computing Decompositions

Having established the theoretical foundation for decomposing forecast densities, the next step is to address how contributions can be computed in practice.

Following Strumbelj and Kononenko (2010) the Shapley value of variable x_j at time t for a model prediction $\hat{f}$ can be written as

$$\varphi_j(\boldsymbol{x}_t;) = \sum_{\boldsymbol{x}_S \subseteq \mathcal{C}(x)\setminus\{j\}} \frac{|\boldsymbol{x}_S|!(n-|\boldsymbol{x}_S|-1)!}{n!} \left[\hat{f}(x_j|\boldsymbol{x}_S \cup \{j\}) - \hat{f}(x_j|\boldsymbol{x}_S)\right], \tag{10}$$

In this equation, $\boldsymbol{x}_S$ denotes a subset of $\boldsymbol{x}$, $|\boldsymbol{x}'|$ denotes its cardinality, $\mathcal{C}(x)\setminus\{j\}$ is the set of all possible variable combinations excluding the j^{th} variable, and $\frac{|\boldsymbol{x}_S|!(n-|\boldsymbol{x}_S|-1)!}{n!}$ is a combinatorial weight. Essentially, Equation 10 describes the weighted sum of the marginal contributions of variable x_j to model prediction $\hat{f}$ in all possible coalitions of variables, where intuitively each of these coalitions and orderings correspond to one path from $\boldsymbol{x}$ to $\boldsymbol{x}'$. In the present context, $\hat{f}$ represents not a scalar prediction but the probability mass in a given bin of the forecast density. The decomposition can then be carried out bin by bin yielding a complete attribution of the density's shape.

A practical challenge is that Equation 10 requires evaluating the model on all possible coalitions of predictors, which is computationally infeasible except in small systems. Strumbelj and Kononenko (2010) propose a Monte Carlo approximation that samples random coalitions. Building on this idea, this paper develops an algorithm tailored to forecast densities (see algorithm 2 in Online Appendix 3). It operates at the level of probability bins, and is designed for real-time monitoring and forecasting applications, regardless of the model used to predict the forecast density.

For one time period t and one variable j, the algorithm evolves as follows:

1. At given time period t, sample a random permutation of variables and collect the variables appearing before variable j in a set, S, and sample a random past time index $t^* \in \{1, \ldots, t-1\}$.
2. Construct a hybrid observation by replacing, in $\boldsymbol{x}_{t^*}$, the variables in S with their values from $\boldsymbol{x}_t$. Create two observation vectors: (i) $\boldsymbol{x}_t^{-j}$, where x_{j,t^*} remains unchanged; and (ii) $\boldsymbol{x}_t^{+j}$, where x_{j,t^*} is replaced by $x_{j,t}$.
3. Evaluate the bin probabilities for both vectors. The difference between the two evaluations yields the marginal contribution of x_j.
4. Repeat the procedure for a number of MC samples, M. The average of these marginal contributions yields an estimate of $\varphi_j^{y^\star}$.

Repeating the algorithm at different time periods and for different variables, x_j, then yields the full sequence of contributions for all variables.

Notably, because the sampling-based algorithm does not explore all possible coalition-time-period combinations, in practice the contributions are not guaranteed to add up to the difference between the forecast density and the historical average. However, the empirical and simulation exercises suggest that this is only a limited issue which can further be mitigated by increasing the number of draws.

In the sampling-based approach of algorithm 2, missing predictors are replaced with draws from their empirical marginals. When predictors are correlated, this may generate off-manifold samples, which has triggered an active debate in the explainable machine-learning literature. Some authors argue that attributions should be *true to the data* by conditioning on the observed dependence structure, while others argue they should be *true to the model*, focusing on how the fitted model reallocates probability mass regardless of empirical correlations (Aas et al., 2021; Chen et al., 2020; Janzing et al., 2020). Which approach is preferable is ultimately application-specific. In the present context, the objective is to decompose how the model generates the forecast density, making marginal sampling the natural choice (Chen et al., 2020).

For completeness, the online appendix outlines a conditional variant of the algorithm and presents simulation studies comparing marginal and conditional approaches. The results suggest that qualitative conclusions remain aligned.

3.2 Simulation Validation and Algorithmic Performance

Before turning to the empirical forecasting application, the behaviour of the proposed models and decomposition algorithms is evaluated in controlled simulation settings. The exercises assess the accuracy, robustness, and interpretability of the framework under known data-generating processes. All implementation details, parameter settings, and numerical results are provided in the online appendix in section 4. The experiments serve three purposes: (i) to examine how the DR and QR models capture different distributional shapes; (ii) to assess the accuracy and convergence of the Shapley-based decomposition algorithm under known data-generating processes; and (iii) to study the robustness of the algorithm across model specifications and data structures.

The first simulation evaluates how the DR and QR models approximate eight distinct data-generating processes, ranging from Gaussian to skewed, heavy-tailed, multimodal, and outlier process designs (Mitchell et al., 2024; Lim et al., 2020; Korobilis and Schröder, 2025b). Each DGP is simulated for $T = \{50, 100, 200, 500\}$, with model fit assessed using

moment-based mean-squared errors and Kullback-Leibler divergence statistics, following Mitchell et al. (2024). Overall, both models perform well and broadly on par, with the DR model achieving slightly lower errors, particularly as sample size increases. The only notable exception is the outlier process, where skewness and kurtosis errors remain relatively large. For the DR model, an additional evaluation assesses its ability to recover the correct number of mixture components. Results indicate that the true component number is accurately identified in all cases except the outlier and highly kurtotic processes, where it is overestimated by one.

The second simulation evaluates the performance of the proposed decomposition algorithms under settings with independent and dependent predictors. Competing methods include Shapley values computed over all coalition-time-period combinations, kernel-based Shapley values, conditional Shapley methods based on k-nearest neighbors (kNN), gradients and integrated-gradients, as well as the sampling-based Shapley approach proposed above. Each algorithm decomposes simulated forecast densities and risk measures (introduced later), with performance assessed using a range of score metrics and correlation statistics. Across both VAR specifications, the Shapley-based methods consistently yield more accurate results, with the sampling-based approach performing on par with the full implementation. Moreover, the resulting decompositions aggregate to the difference between the forecast and background distributions with negligible error. The results suggest that the proposed approach provides reliable and interpretable estimates of variable contributions.

The final simulation examines whether the proposed decomposition algorithm performs consistently across both the DR and QR models. Across sample sizes, both frameworks yield accurate decompositions, with the DR model showing slightly larger error and score metrics but more favourable correlation statistics than the QR. As sample size increases, performance converges. Overall, the results demonstrate that the proposed algorithm operates robustly across modelling approaches and sample sizes, confirming its suitability for a wide range of empirical applications.

4 Empirical Exercises

As an application of the proposed methods, this section analyzes forecast densities for key macroeconomic aggregates, focusing on U.S. inflation during the pandemic period. This episode provides a timely and policy-relevant case for illustrating how the framework can decompose and interpret risks around forecasts through the lens of the

models. A large and growing literature has examined the drivers of the pandemic-era inflation surge, typically focusing on realized inflation outcomes. Ball et al. (2022) stress the role of labor-market tightness and headline pass-through, while Blanchard and Bernanke (2023) emphasize early supply shocks, labor-market persistence, and the stance of monetary policy. Ha et al. (2021) and Hall et al. (2023) highlight global demand, commodity prices, fiscal and monetary responses, and supply-chain pressures, with the latter also documenting international spillovers. More narrowly, LaBelle and Santacreu (2022) and Abbai et al. (2022) attribute a substantial share of U.S. inflation to global supply disruptions, while Bai et al. (2024) quantify the causal effects of shipping congestion. Giannone and Primiceri (2024) argue that demand forces dominated, whereas Cascaldi-Garcia et al. (2024) emphasize synchronized global drivers, including labor-market tightness, commodity shocks, and the Ukraine war. The present paper revisits this episode from a different angle: instead of realized inflation, it analyzes forecast densities and their decomposition, thereby illustrating how the proposed framework can shed light on the sources of inflation risks rather than levels alone.

The application proceeds in three steps: first, the dataset and forecast design are introduced; second, forecast densities are evaluated using standard metrics; and third, the decomposition framework is applied to trace the predictors behind changes in distributions and risk measures.

4.1 Data

The baseline dataset spans 1974Q2-2024Q4 and uses as the primary target the annualized quarterly change in U.S. PCE inflation excluding food and energy, the Federal Reserve's preferred gauge of underlying inflation pressures. Additional results for headline PCE, CPI inflation, and CPI core inflation are reported in the online appendix. The dataset, summarized in Table 1, is deliberately compact but covers the main channels emphasized in the literature: domestic and global activity, commodity prices, inflation expectations, financial conditions, and monetary policy.

As such, it is consistent with earlier work that links inflation risks to a small set of forward-looking predictors. In particular, Lopez-Salido and Loria (2024) highlight the role of inflation expectations and financial spreads. Reflecting this, the present dataset incorporates one-year-ahead and long-run expectations, as well as the excess bond premium of Gilchrist and Zakrajšek (2012)[6], a measure of financial conditions within the

[6] Gilchrist and Zakrajšek (2012) find the EBP to be particularly informative about future economic activity. In comparison to broad financial condition indices such as the NFCI, the EPB is constructed

broader class of indicators often found to be an important predictor of macroeconomic uncertainty (Banerjee et al., 2024; Korobilis et al., 2021; Lopez-Salido and Loria, 2024; Tagliabracci, 2020). Finally, domestic activity is proxied by an activity factor extracted from the FRED-QD database of McCracken and Ng (2020), spanning output and activity proxies, labor market and financial indicators.[7]

Table 1: Data Overview

Macroeconomic Indicators	Source	Mnemonics	Tcode[1]
Personal Consumption Expenditure excl. Energy & Food	FRED	PCEx	4
U.S. business cycle indicator	FRED-QD[2]	BC_d	1
Excess Bond Premium	Gilchrist and Zakrajšek (2012)[3]	EBP	1
Federal Funds Rate	FRED	FFR	1
OECD+6 industrial production	Baumeister and Hamilton (2019)[4]	BC_g	4
Energy price index	"Pink Sheet", The World Bank[5]	Energy	4
Food price index	"Pink Sheet", The World Bank[5]	Food	4
Metal price index	"Pink Sheet", The World Bank[5]	Metals	4
1 year ahead inflation expectations	Survey of professional forecasters[6]	EXP_{1Y}	1
Long-run inflation expectations	FRB/US project database[7]	EXP_{LR}	1

[1] Tcode identifies the time series transformation: level (1), log-differences (4).
[2] The data is taken form the FRED-QD database. The indicator is based on own computations.
[3] https://sites.google.com/site/cjsbaumeister/datasets?authuser=0.
[4] https://www.worldbank.org/en/research/commodity-markets.
[5] https://www.federalreserve.gov/econres/notes/feds-notes/ebp_csv.csv.
[6] https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/survey-of-professional-forecasters.
[7] https://www.federalreserve.gov/econres/us-models-about.htm.

At the same time, the dataset broadens the scope by systematically incorporating global business cycle dynamics, proxied by the OECD+6 industrial production index of Baumeister and Hamilton (2019), and commodity price indices for energy, food, and metals, intended to capture global demand and supply-side channels that have been central to the post-pandemic inflation literature (Abbai et al., 2022; Bai et al., 2024; Ball et al., 2022; Blanchard and Bernanke, 2023; Cascaldi-Garcia et al., 2024; Giannone and Primiceri, 2024; Ha et al., 2021; Hall et al., 2023; LaBelle and Santacreu, 2022).

The final indicator is the federal funds rate (FFR), included as the standard measure of interest rates and monetary policy. Yet, because interest rates reflect not only central bank decisions but also endogenous macro-financial forces, relying on the raw FFR can generate price or output puzzles in applied research. To isolate the policy-driven component,

entirely from financial market data and captures information not already contained in macroeconomic variables (Reichlin et al., 2020).

[7] The database spans a large set of economic indicators that cover 14 broad groups - NIPA; industrial production; employment & unemployment; housing; inventories, orders & sales; prices; earnings & productivity; interest rates; money & credit; household balance sheets; exchange rates; others (uncertainty & expectations); stock markets; non-household balance sheets. Prior to extraction, broad aggregates are excluded as recommended by Stock and Watson (2012), and the data is transformed to ensure stationarity. The activity factor is then extracted using partial least squares for one-step-ahead inflation.

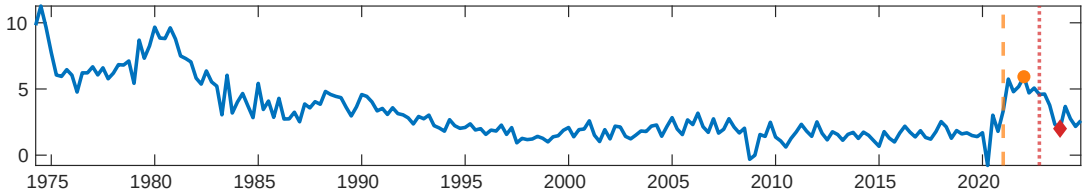
this paper also constructs a monetary policy proxy from high-frequency information, re-estimating the Proxy-SVAR of Miranda-Agrippino (2016) on an updated sample. The resulting structural shocks are employed in the density decomposition exercises, while the raw FFR is retained for the forecasting exercise.

All predictors are standardized before estimation. Apart from these transformations, the data remains unaltered and is not adjusted for pandemic outliers.

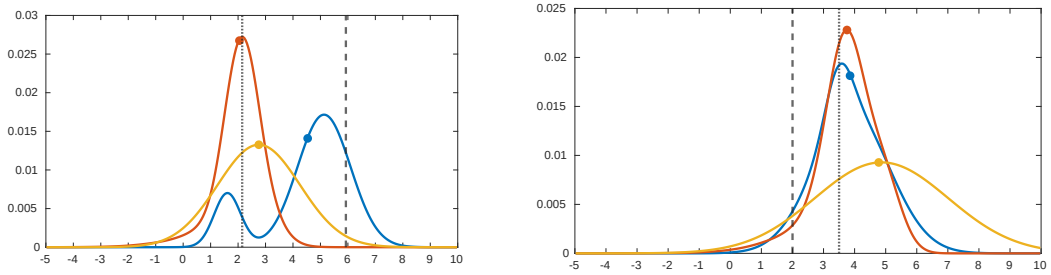
4.2 Forecasting Exercise

Figure 1: examples of 4-step ahead forecast distributions

(a) Realized inflation rate



(b) 2022Q1 (c) 2023Q4



Notes: Top panel: Realized PCE excluding energy and food inflation rate. The dashed (dotted) vertical line denotes the vintage 2021Q1 (2022Q4). The circle (diamond) denotes the corresponding 4-quarter ahead forecast target, i.e. 2022Q1 (2023Q4). Bottom panels: The corresponding forecast densities estimated with the density regression model (blue), the quantile regression model (red), and the UC-SV model (yellow). All models are estimated as specified in subsection 2.4. The dots on the density correspond to the model specific point estimate. The gray dashed line indicates the ex-post realized value the dotted line indicates the FOMC forecast, constructed from the summary of projections.

Figure 1 provides an illustrative example of the data and associated forecast densities from the density regression and quantile regression models, alongside the UC-SV benchmark of Stock and Watson (2007). The top panel plots U.S. PCE inflation excluding food and energy, overlaid with two vintages used to generate forecasts. The bottom panels display the corresponding forecast densities for two key moments: the peak of the high-inflation episode and the point in 2023 when inflation returned to target.

In the first case, both the UC-SV and the QR densities are centred near the FOMC forecast and the 2% target, though the UC-SV implies generally large inflation uncertainty.

By contrast, the DR density is bimodal, with one small mode around the target but most probability mass close to the outcome realized a year later. For the second vintage, both the DR and QR densities are centred near the FOMC forecast and point to higher inflation than observed ex-post, with some right skew, while the UC-SV again indicates pervasive uncertainty centred at historically high levels.

The examples illustrate that the flexible DR and QR models yield richer insights into inflation risks, especially during critical periods such as the post-pandemic surge and the subsequent disinflation, whereas the UC-SV benchmark cannot capture distinct modes or distributional asymmetries. To assess whether these qualitative differences translate into systematic forecasting gains, this section evaluates model performance over the full sample, based on calibration tests and score-based metrics. For all models, estimation begins with 50% of the total sample. Forecasts are then generated recursively: at each vintage the models are re-estimated, direct forecasts are produced for horizons from 1 to 12 quarters ahead, and the sample is extended by one observation.[8]

Calibration is first assessed using probability integral transforms (PITs), which measure the probability mass the model assigns to the realized outcome, $\text{pit}_{t+h|t} = \int^{y_{t+h}} -\infty f_x(y)_{t+h|t} dy$. Under correct specification, PITs should be uniformly distributed and, for one-step-ahead forecasts ($h = 1$), serially independent. Rossi and Sekhposyan (2019) provide formal tests of these properties, applied here to the overall distribution as well as separately to the centre and tails. The results are summarized in the top panel of Table 2.

The calibration tests suggest that the flexible models are more consistent with the true shape of the conditional forecast distribution than the UC-SV benchmark. For DR, correct calibration cannot be rejected at most horizons, with only isolated rejections in the left tail. QR exhibits more frequent miscalibration, particularly at short horizons ($h \leq 4$). By contrast, UC-SV fails calibration tests almost uniformly, with the exception of the right tail at $h = 8, 12$. Overall, the evidence suggests that semi-parametric models, especially DR, yield density forecasts that better capture asymmetries and tail behaviour.

While calibration speaks to statistical consistency, it does not measure forecast accuracy across the distribution. To complement the PIT-based diagnostics, forecast densities are also evaluated using scoring rules: quantile scores (QS) following Manzan (2015), which evaluate predictive accuracy at specific quantiles, and the continuous

[8]Please note that for $h > 1$, the direct multi-step approach inherently produces serially correlated and overlapping forecast errors, which can lead to misspecification (see, e.g., Koop, 2013). Whether variational Bayes is more robust to this issue than Gibbs sampling remains an open question. To implement specification tests in the presence of serial correlation, such as with multi-step-ahead density forecasts, Rossi and Sekhposyan (2019) develop bootstrap procedures.

Table 2: Density specification tests - PCE excluding food and energy inflation

	Density specification tests															
	full		center		left tail		right tail		full		center		left tail		right tail	
	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM
	$H = 1$								$H = 2$							
DR	0.96	0.20	0.96	0.20	0.93	0.31	0.82	0.11	0.56	0.04	0.56	0.05	**0.49**	0.05	0.20	0.01
QR	**1.39**	0.40	**1.39**	0.61	**1.08**	**0.38**	0.35	0.03	**1.60**	**0.70**	1.27	**0.78**	0.35	0.03	**1.60**	**1.22**
UC-SV	**2.96**	**2.37**	**2.96**	**3.20**	**2.25**	**1.62**	**2.00**	**1.48**	**2.35**	**1.98**	**2.35**	**2.60**	**2.14**	**1.52**	**1.63**	**1.22**
	$H = 3$								$H = 4$							
DR	1.12	0.21	1.12	0.38	**0.51**	0.04	0.38	0.03	1.08	0.18	1.08	0.29	**0.67**	**0.13**	0.36	0.02
QR	**2.20**	**2.06**	**2.20**	**2.82**	**1.43**	**0.59**	**1.87**	**2.01**	**2.13**	**1.92**	**2.13**	**2.41**	**0.61**	**0.10**	**2.10**	**2.75**
UC-SV	**2.88**	**2.25**	**2.88**	**3.35**	**2.04**	**1.47**	**1.74**	**0.87**	**2.68**	**1.93**	**2.68**	**2.68**	**2.14**	**1.48**	**1.69**	**0.92**
	$H = 8$								$H = 12$							
DR	1.04	0.14	0.45	0.04	**1.04**	**0.36**	0.59	0.12	1.05	0.59	1.05	0.65	**0.99**	**0.33**	1.04	0.71
QR	1.25	0.38	1.03	0.21	**0.58**	**0.08**	**1.25**	**1.03**	1.01	0.18	0.86	0.21	0.47	0.06	1.01	0.27
UC-SV	**2.25**	**1.39**	**2.25**	**1.86**	**2.10**	**1.47**	1.08	0.36	**2.28**	**1.30**	**2.28**	**1.83**	**2.04**	**1.40**	0.92	0.16

	Quantile scores							CRPS	*Quantile scores*							CRPS
	0.05	0.1	0.2	0.5	0.8	0.9	0.95		0.5	0.1	0.2	0.5	0.8	0.9	0.95	
	$H = 1$								$H = 2$							
DR	-0.10	-0.15	-0.22	-0.31	-0.25	-0.20	-0.15	0.46	-0.15	-0.20	-0.27	-0.36	-0.29	-0.22	-0.16	0.53
QR	-0.11	-0.14	-0.20	-0.30	-0.23	-0.16	-0.11	0.42	-0.11	-0.16	-0.23	-0.34	-0.28	-0.18	-0.13	0.50
UC-SV	-0.15	-0.24	-0.33	-0.28	-0.32	-0.24	-0.15	0.54	-0.15	-0.24	-0.34	-0.30	-0.32	-0.24	-0.15	0.57
	$H = 3$								$H = 4$							
DR	-0.16	-0.22	-0.29	-0.39	-0.32	-0.24	-0.18	0.57	-0.21	-0.26	-0.33	-0.40	-0.31	-0.23	-0.17	0.59
QR	-0.11	-0.15	-0.24	-0.35	-0.35	-0.30	-0.23	0.56	-0.11	-0.16	-0.23	-0.39	-0.38	-0.31	-0.25	0.59
UC-SV	-0.16	-0.25	-0.35	-0.30	-0.35	-0.25	-0.15	0.59	-0.16	-0.25	-0.36	-0.34	-0.36	-0.27	-0.17	0.62
	$H = 8$								$H = 12$							
DR	-0.26	-0.34	-0.44	-0.55	-0.47	-0.38	-0.28	0.84	-0.17	-0.24	-0.40	-0.55	-0.50	-0.41	-0.33	0.83
QR	-0.11	-0.17	-0.24	-0.42	-0.52	-0.34	-0.26	0.66	-0.11	-0.17	-0.23	-0.48	-0.53	-0.42	-0.30	0.72
UC-SV	-0.16	-0.26	-0.38	-0.44	-0.42	-0.31	-0.20	0.73	-0.16	-0.26	-0.38	-0.46	-0.45	-0.32	-0.19	0.76

The top half of the table contains the test statistics for the density specification tests in Rossi and Sekhposyan (2019). KS denotes the Kolmogorov-Smirnov and CVM the Kramér-von Mises based test statistic. Bold values indicate that the null hypothesis of correct specification can be rejected at 5% significance level. The bottom part of the table contains quantile scores following Manzan (2015), where smaller values indicate better performance and continuous ranked probability scores (CRPS).

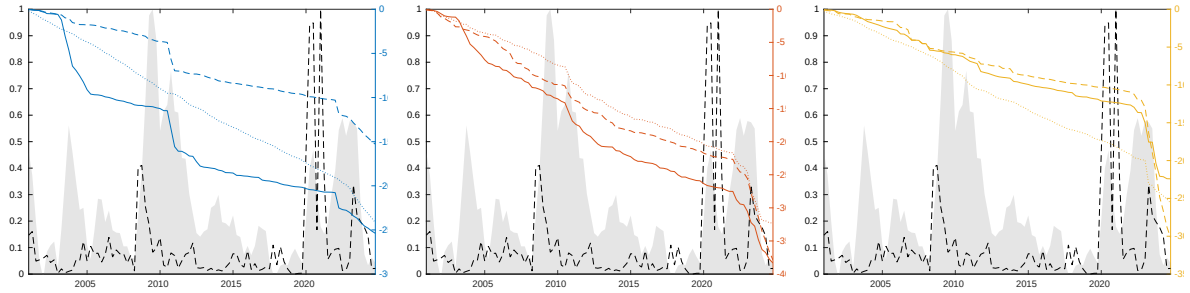
ranked probability score (CRPS), which summarizes overall density fit. The corresponding results are reported in the bottom panel of Table 2.

Here, the picture is more nuanced. While from the perspective of CRPS the DR and QR outperform the UC-SV for $h \leq 4$, the UC-SV has the lower quantile scores in the majority of cases for shorter horizons. In turn, for longer horizons, particularly the DR model improves upon the UC-SV. To put these results into perspective, the online appendix tests for equal accuracy of QS and CRPS between the competing models. For the DR, differences remain statistically insignificant for $h > 1$ and for the QR the UC-SV performs significantly better only in the minority of cases.

Taken together, the results indicate that while the flexible models offer clear improvements in calibration, the evidence for systematic gains in forecast accuracy is more local. This raises the question of whether improvements emerge in particular episodes, rather than uniformly across the sample. To investigate, Figure 2 plots cumulative quantile scores at the four-quarter horizon ($h = 4$), the main horizon of

interest for the subsequent analysis of inflation risks, for $q = \{0.10, 0.50, 0.90\}$.

Figure 2: Cumulative quantile scores for $h = 4$



(a) 10^{th} percentile (b) 50^{th} percentile (c) 90^{th} percentile

Notes: The cumulative quantile scores for the DR, QR, and UC-SV model are indicated by the solid, dashed, and dotted line respectively. The black dashed line indicates the centred moving variance of the underlying PCE excluding energy and food inflation rate. The gray shaded area indicates the interquartile range of the survey of professional forecasters (SPF) for the 4-step ahead qoq core PCE inflation rate. Early observations were extended backwards using the interquartile range of the CPI.

The panels show that in all cases the more flexible variants deliver meaningful performance gains precisely when forecast disagreement among SPF forecasters is high, as indicated by the shaded areas. By contrast, these gains do not systematically coincide with realized volatility of inflation, implying dynamics beyond those captured by the UC-SV model. This pattern suggests that flexible models are particularly valuable as a complement to simpler models in periods when forecasting is most challenging, such as during the recent high-inflation episode.[9]

Corresponding results for headline PCE and CPI, as well as core CPI inflation, are reported in the online appendix and broadly consistent with the findings presented here. The online appendix also provides point forecast estimates and corresponding Diebold-Mariano test statistics for all inflation measures. Overall, the UC-SV model tends to deliver more precise point forecasts for core inflation, while the flexible methods perform better for headline inflation. However, performance differences are statistically significant only in selected cases.

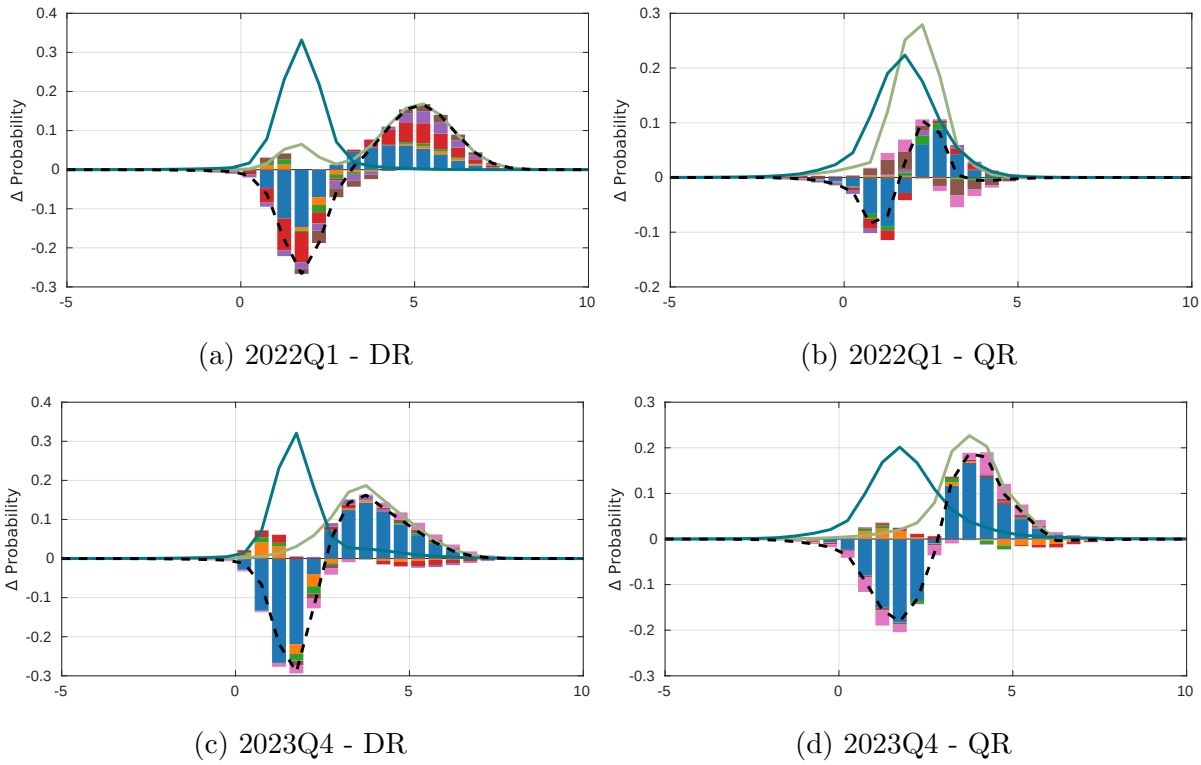
4.3 Forecast Distribution Decompositions

The forecasting results above showed that the flexible DR and QR models generate well-calibrated and informative forecast densities, particularly during periods of heightened

[9] Figure 3 in the online appendix shows that differences between DR and QR estimated quantile levels arise mainly in high-inflation periods, with QR tending to predict lower quantiles particularly at the median and 10th percentile.

uncertainty and forecast disagreement. In this section, the estimated densities are used to illustrate how the proposed Shapley-based decomposition algorithm can uncover how the fitted models, through their structure and estimated parameters, reallocate probability mass when economic conditions change.

Figure 3: Decomposition of the Forecast Distribution for $h = 4$



(a) 2022Q1 - DR

(b) 2022Q1 - QR

(c) 2023Q4 - DR

(d) 2023Q4 - QR

Notes: The teal density indicates the average historical density. The olive density indicates the forecast density for 4-step ahead annualized quarterly PCE excluding energy and food inflation. The dashed black line indicates the difference between the predicted density and the historical average. The left column presents estimates computed with the density regression (DR) model and the right column results for the quantile regression (QR) model. The stacked bars contain the contributions, with: ● dynamic persistence, ● domestic business cycle, ● global business cycle, ● commodity prices, ● financial conditions, ● monetary policy, ● inflation expectations.

The algorithm, introduced in Section 3.1, assigns contributions to individual predictors based on their marginal effect on the shape and location of the forecast density. The resulting attributions provide a reduced-form, model-consistent view of the drivers of forecast distributions - a "true-to-the-model" perspective in the sense discussed in Section 3.1. The exercise is therefore intended to demonstrate the mechanics and interpretive value of the decomposition method, rather than to identify structural relationships.[10]

[10]While the decompositions are reduced-form, in principle causal interpretations can be made for predictors that correspond to externally identified shocks. In particular, for monetary policy, the analysis

Figure 3 presents the results for the 4-quarter ahead forecasts densities for the same two representative forecast vintages introduced above, i.e. the peak of the high-inflation period (2021Q1) and the return to target (2023Q4), chosen to illustrate how contributions to inflation risk evolved across distinct macroeconomic environments. The decomposition algorithm is identical for both the DR and QR model and applied in an out-of-sample pseudo-real-time fashion using algorithm 2, based on $M = 1000$ random coalition-time-samples. In each panel, the teal-coloured density corresponds to the background data density (i.e., the historical average forecast density), the olive-coloured density to the model's forecast density, and the dashed line to their difference. The bars represent the decomposition, which sums to this difference. For clarity of presentation, the constant term, lagged variables, commodity prices, and inflation expectations are grouped into their respective categories.

Starting with the first vintage, the DR model's right-shifted second mode in the forecast density, as observed earlier, primarily reflects the contributions of commodity prices and, to a lesser extent, past inflation levels.[11] Smaller contributions stem from financial conditions and the carry-over effects of loose monetary policy during and after the pandemic. Business cycle factors remain relatively muted, primarily redistributing probability mass toward lower inflation outcomes. Taken together, these contributions can help explain the DR model's heavier upper tail and the emergence of a secondary mode at elevated inflation rates, though differences in how the two models capture non-linearities, parameter uncertainty, or temporal heterogeneity may also play a role.

In contrast, for the QR model, while commodity prices and inflation dynamics still shift probability mass toward higher realizations, this redistribution is considerably less pronounced and concentrated around more moderate inflation outcomes. In addition, monetary policy and inflation expectations contribute to shifting mass toward lower outcomes, resulting overall in a more symmetric, unimodal density centered near the 2% target. This pattern may suggest that the QR-based densities place less weight on the persistence of past inflation and pandemic-related shocks, yielding a forecast that appears less skewed and more tightly anchored around the target.

While for the first vintage the two models offered rather different interpretations of the real-time data flow, in the second vintage, corresponding to the disinflation phase, the forecast densities appear driven by broadly similar forces across models. The largest

includes high-frequency monetary policy surprises as discussed in subsection 4.1, allowing for a structural interpretation of their contributions. For all other variables, the attributions remain reduced-form.

[11] Importantly, rather than capturing elevated inflation alone, the bars also reflect indirect effects operating through other variables.

contributions originate from elevated inflation and inflation inertia in preceding quarters. Commodity prices and business cycle factors primarily shift probability mass from above-target outcomes toward levels consistent with the inflation objective. However, inflation expectations now contribute to forecasts of higher inflation outcomes, suggesting that expectations played a more prominent role in sustaining near-term inflation risks during the return-to-target phase through the lens of both models.

Taken together, the two vintages highlight how contributions to inflation risk can differ markedly across macroeconomic environments and between models. During the inflation peak, forecast densities reflected a broader set of upward pressures and greater asymmetry. Although the DR and QR models differ in how they allocate these contributions, both point to a broad normalization of inflation risks as price pressures eased, with the remaining risk of high inflation outcomes predominantly associated with inflation inertia, which could in part reflect the delayed pass-through of past shocks.

5 From Density Decompositions to Risk Measures

Building on the density decompositions above, this section summarizes the information contained in the full predictive distributions through risk measures that track the evolution of inflation risks over time and offer a compact and interpretable representation of the model's forecasts. By extending the decomposition framework to transformations of forecast densities, such as risk measures, the analysis enables both a direct comparison with the broader literature on post-pandemic inflation dynamics and a dynamic cross-check of the decomposition results, showing that the factors contributing to shifts in forecast densities also shaped the changing balance of inflation risks during the post-pandemic period.

5.1 Risk Measures for Flexible Density Forecasts

The macroeconomic uncertainty literature often summarizes forecast distributions through quantile-based risk measures. While convenient, such statistics focus on limited regions of the distribution, typically one tail or its centre, and can therefore miss important shifts elsewhere: Quantiles or the Value-at-Risk (VaR) shift only when probability mass crosses their percentile threshold; redistributions entirely above or below leave them unchanged. The interquartile range (IQR) depends solely on the 25th and 75th percentiles and is invariant to changes outside this central 50% region, so heavy- and light-tailed distributions can share identical IQRs. Expected Shortfall (ES)

and its upper-tail counterpart, the Expected Longrise (LR), react only to movements of probability mass within their respective tails, remaining fixed when the rest of the distribution shifts. Together, these invariance properties show that quantile- and tail-mean-based measures capture only partial aspects of distributional change, motivating the use of statistics that reflect adjustments across the entire support of the predictive density.

From a purely statistical standpoint, two forecast distributions with the same mean may appear equally desirable, even if one places far more probability mass in the tails. Yet for a policymaker concerned with price stability, these distributions differ substantially: a central bank would generally prefer the one narrowly centred on target to another with fatter tails, reflecting aversion to extreme outcomes in either direction. In practice, central banks are hence not risk-neutral and attach asymmetric costs to deviations from target - a 10-percentage-point overshoot is disproportionately more undesirable than a 1-percentage-point overshoot (Bernanke, 2003). A comprehensive risk measure should therefore go beyond summarizing statistical characteristics and reflect the pricing of risk implied by policy preferences (Kilian and Manganelli, 2007, 2008).

To incorporate such preferences, Kilian and Manganelli (2007, 2008) propose risk measures grounded in the risk-management and finance literature (e.g., Machina and Rothschild, 1987). In this framework, upside risk corresponds to inflation outcomes, denoted by π, exceeding an upper threshold $\bar{\pi}$, and downside risk to π falling short of a lower bound $\underline{\pi}$, where $\underline{\pi} \leq \bar{\pi}$. For point targets, such as the Federal Reserve's inflation target, $\underline{\pi} = \bar{\pi} = \pi^*$, and preference parameters α and β, the measures are defined as

$$DR_\alpha \equiv -\int_{-\infty}^{\underline{\pi}} (\underline{\pi} - \pi)^\alpha dF_\pi(\pi), \text{ with } \alpha \geq 0 \tag{11}$$

$$EIR_\beta \equiv \int_{\bar{\pi}}^{\infty} (\pi - \bar{\pi})^\beta dF_\pi(\pi), \text{ with } \beta \geq 0, \tag{12}$$

Here, DR_α and EIR_β denote deflation and excess-inflation risk, combining probability and preference weightings of deviations from target while remaining consistent with the predictive distribution.

For $\alpha = \beta = 0$, the measures reduce to the marginal probabilities of inflation undershooting or overshooting the target, summarizing the likelihood of threshold breaches irrespective of their size. Setting $\alpha = \beta = 1$ yields a formulation related to expected-shortfall-type measures, where risk is proportional to the average magnitude of deviations beyond the target. When $\alpha = \beta = 2$, the measures reflect a semi-variance-type concept that penalizes larger deviations disproportionately more

than smaller ones. Increasing α or β thus raises the sensitivity of the measure to the size of the deviation, reflecting stronger aversion to extreme inflation outcomes.

Beyond their tractable form, Kilian and Manganelli (2008) show that a risk-managing central bank can be represented as an expected-utility-maximizing central bank with a suitable loss function,

$$L = w \cdot \mathbb{I}(\pi < \underline{\pi})(\underline{\pi} - \pi)^{\alpha} + (1 - w) \cdot \mathbb{I}(\pi > \bar{\pi})(\pi - \bar{\pi})^{\beta}, \tag{13}$$

where $w \in [0, 1]$ governs the relative weight assigned to downside versus upside risks. Taking expectations yields the balance of risk or expected loss,

$$BR_{\alpha,\beta} \equiv E(L) = -wDR_{\alpha} + (1 - w)EIR_{\beta}. \tag{14}$$

This formulation directly links the empirical risk measures to the standard decision-theoretic framework of monetary policy, allowing consistent interpretation across preference specifications, including non-integer or regime-dependent values of α and β, and providing a bridge between structural policy models and empirical forecast-based risk assessments.

5.2 Historical Inflation Risk Measures

The baseline specification adopts $\alpha = \beta = 2$, representing a risk-averse policymaker concerned with limiting both deflation and excessive-inflation risks, consistent with the notion of "low and stable inflation" (Blinder, 1997; Kilian and Manganelli, 2007, 2008; Svensson, 2002). To align with the Federal Reserve's inflation objective of 2 percent annually, the bounds are set to $\underline{\pi} = \bar{\pi} = 2$, and equal weights are applied to upside and downside risks ($w = 0.5$). The risk measures in Equation 11 and Equation 12 are computed from the predictive distributions constructed in the preceding sections, with integrals evaluated numerically using Gaussian quadrature methods. This yields a efficient procedure for translating the full forecast densities into preference-weighted measures of inflation risk and extends naturally to other density forecasting models.[12] [13]

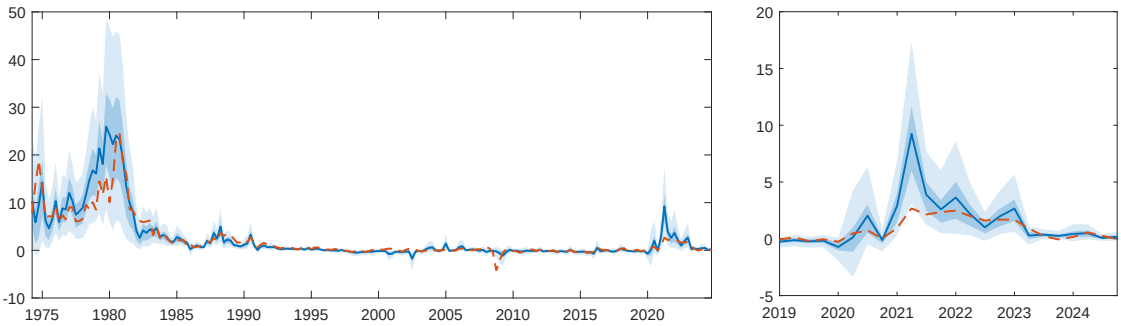
[12] Alternative parametrizations could in principle capture different policy regimes. A lower weight on upside risk ($w < 0.5$ or $\beta < \alpha$) might reflect caution about tightening during a fragile recovery, whereas greater concern about deflation when rates are near the effective lower bound could correspond to $w > 0.5$.

[13] As in Kilian and Manganelli (2007, 2008), the approach does not rely on an elicitable loss function. Estimation and evaluation criteria are hence not formally aligned (Patton et al., 2019). This limitation is common in the literature and reflects the broader challenge of constructing elicitable objectives for distributional forecasts.

Figure 4 displays the estimated balance of risk for four-quarter-ahead inflation. The left panel shows the full sample for completeness and historical context, noting that the Federal Reserve operated without an explicit inflation target before 2012 and only implicitly from the mid-1990s onward. The right panel zooms in on the post-pandemic high-inflation period. The solid blue line corresponds to the DR model and the dashed red line to the QR model. Positive (negative) values indicate periods of elevated excess-inflation (deflation) risk.

Figure 4: The Balance of Risk

(a) BR for $\alpha = \beta = 2$ and $h = 4$ (b) BR 2019Q1 to 2024Q4



Notes: The solid blue line indicates the estimates for the density regression model over the full sample obtained. The dark shaded area corresponds to the 50% and the light shaded area to the 90% credible interval. The dashed red line shows the corresponding estimates computed with the QR model.

With these preference parameters, both models show elevated inflation risks through the early 1980s, reflecting the culmination of the Great Inflation. The QR model registers the buildup about a year later, around 1979, than the DR model, suggesting a lagged response in quantile-based forecasts. Inflation risks then declined steadily into the early 1990s alongside the Volcker-era monetary tightening, with two temporary reversals: one around 1987, coinciding with the "1987 inflation scare" Goodfriend (1993), and another during the 1990-91 Gulf War. Over this phase, projections from both models remain broadly aligned. From the mid-1990s to 2010, inflation risks are generally subdued, interrupted by a small uptick around 2005 coinciding with a series of oil price increases. Troughs follow the 2001 dot-com recession and align with the 2008-09 global financial crisis, during which the QR model assigns stronger downside risks and indicates their increase slightly earlier than the DR model. Thereafter, both measures remain flat until the onset of the COVID-19 pandemic, when four-quarter-ahead risks rise sharply to their highest levels since the early 1980s before subsiding from mid-2021 onward. While turning points and overall dynamics are closely aligned, the QR model generally reports lower inflation risks throughout this sample;

however, these deviations are minor and remain well within the DR model's 90%, and often even 50%, credible intervals. Online appendix figures show consistent results for $h = 1$ and for alternative inflation measures, with peaks and troughs more pronounced for headline inflation and shorter forecast horizons.

Overall, both the DR and QR models deliver economically sensible and closely aligned assessments of inflation risk, responding plausibly to changing macroeconomic conditions. Their convergence across different modeling approaches confirms that the resulting risk measures provide a coherent and reliable summary of the evolving balance of inflation risks, forming a sound empirical basis for the subsequent analysis.

5.3 Decomposing Risk Measures

Having established the overall evolution of inflation risks, this analysis extends the decomposition framework introduced earlier to the entire post-COVID inflation episode. Rather than focusing on selected forecast vintages, this exercise applies the Shapley-based algorithm to the full sequence of risk measures, tracing how predictor contributions evolve as inflation pressures build and subsequently recede. Because the risk measures incorporate policy preferences, the resulting decompositions are likewise preference-consistent, providing a dynamic and policy-relevant account of the forces shaping inflation risks over time. Analogously to Section 3, the decompositions are defined in terms of Shapley values that attribute deviations of each risk measure from its historical mean to individual predictors, as follows:

Definition 5.1 (*Contributions to Risk Measures*)

A predictor of downside risk at time t for horizon $t+h$ is a variable, x_j, for which

$$\varphi_{t,j}^{DR_\alpha} = DR_{\alpha,t+h|t} - \mathbb{E}\left(DR_{\alpha,t+h}\right) - \sum_{i\in[1,\ldots,n]:\neq j} \varphi_i^{DR_\alpha} < 0.$$

a predictor of upside risk is a variable, j, for which

$$\varphi_{t,j}^{EIR_\beta} = EIR_{\beta,t+h|t} - \mathbb{E}\left(EIR_{\beta,t+h}\right) - \sum_{i\in[1,\ldots,n]:\neq j} \varphi_i^{EIR_\beta} > 0.$$

If $|\varphi_j| > |\varphi_i|$, for $i \neq j$, then variable x_j is a bigger risk factor at time t than variable x_i. The definition for the balance of risks, BR, is analogous.

A contributor to downside (upside) risk is thus a variable that contributes negatively (positively) to deviations of the corresponding risk measure from its historical average.

The underlying algorithm is identical to that used for density decompositions in Section 3, despite the risk measures being non-linear transformations of the predictive density. It is in full provided by algorithm 3 in the online appendix.

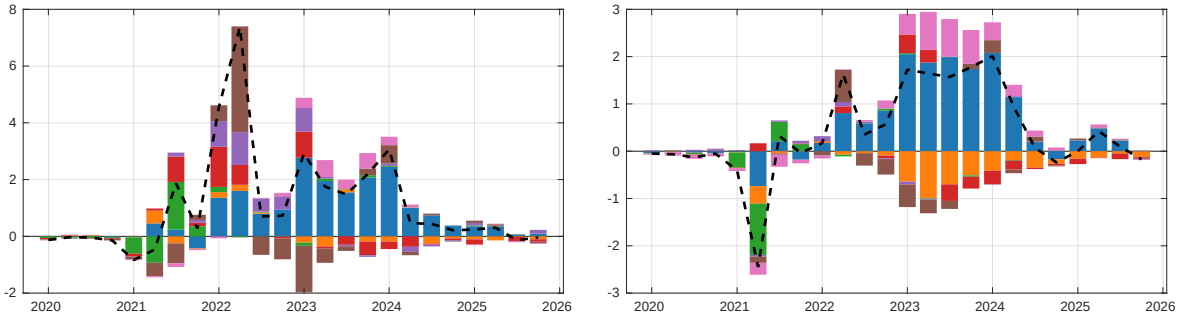
However, two additional considerations arise in this context. First, because the risk measures depend on the preference parameters α and β, the identification and relative magnitude of the contributions are inherently preference-specific. For instance, if the central bank places weight solely on downside risks ($w = 1$), a rise in commodity prices that shifts probability mass only within the upper tail, raising the likelihood of extreme inflation, should not appear as a relevant predictor of risks. More generally, a predictor's contribution increases with the degree of risk aversion and with how far away from the target it shifts probability mass. Hence, just as the risk measures themselves, the risk decompositions must be interpreted conditional on the underlying preference structure, consistent with the principle that any meaningful risk measure must reflect the preferences of the economic agent (Machina and Rothschild, 1987).

Second, when the forecasted risk measure persistently lies below its historical average, the resulting Shapley decomposition produces a negative "overhang", reflecting that the aggregate contribution of predictors offsets the downward deviation of the forecast from the mean. This property follows directly from the efficiency of the decomposition and ensures that contributions remain additive and interpretable even when risks fall below their long-run levels. A simple and effective adjustment is to restrict the background sample used to compute the decompositions, thereby anchoring the decompositions to an economically meaningful benchmark such as the Great Moderation. In the present context, doing so centres the reference point on a period of stable inflation dynamics, making positive and negative contributions more interpretable relative to a well-understood regime of price stability.[14]

Figure 5 summarizes the dynamic contributions of key predictors to the balance of inflation risks over the post-COVID period. The panels show how inflation persistence, commodity prices, inflation expectations, financial and real-economic conditions, and monetary policy alternately amplified or mitigated excess-inflation risks as inflation pressures evolved, for both the DR and QR models. The horizontal axis indexes the forecast target date, with forecasts computed in pseudo-real time four quarters prior.

[14] Alternatively, conditional sampling, as discussed in the online appendix, can be employed to define the benchmark expectation by automatically selecting relevant benchmark samples. However, the resulting benchmarks may be less intuitive to interpret.

Figure 5: Decomposition of the balance of risk $h = 4$



(a) DR (b) QR

Notes: The dashed black line indicates the difference between the predicted balance of risk 4-quarters ahead and as well as the historical average of the 4-quarter ahead balance of risk for PCE excluding energy and food inflation. The left column presents estimates computed with the density regression (DR) model and the right column results for the quantile regression (QR) model. The stacked bars contain the contributions, with: ● *dynamic persistence,* ● *domestic business cycle,* ● *global business cycle,* ● *commodity prices,* ● *financial conditions,* ● *monetary policy,* ● *inflation expectations.*

The chart indicates that, through the lens of both models, the post-pandemic inflation episode can be divided into four distinct phases. The first phase, spanning the target quarters in 2021, is characterized mostly by strong contributions from foreign and domestic business-cycle factors. These mirror the dynamics of the COVID-19 recession and subsequent recovery, initially exerting downward pressure on inflation risk before turning sharply positive as demand rebounded. This pattern aligns with the findings of Cascaldi-Garcia et al. (2024), Ha et al. (2021), Blanchard and Bernanke (2023), and Giannone and Primiceri (2024), who similarly attribute much of the post-pandemic inflation surge to demand-side forces and global activity. The timing of the contributions in this exercise appears slightly earlier than in Giannone and Primiceri (2024), suggesting that the risk-based approach captures evolving pressures in the predictive distribution before they fully materialize in realized inflation.

The second phase covers the first half of 2022, corresponding to forecasts made in early 2021. Inflation risks during this period, most prominently in the DR model, are largely associated with elevated commodity prices, financial conditions, and monetary policy. The commodity price channel possibly reflects the well-documented supply-side pressures from global supply chain disruptions and energy market shocks, consistent with Abbai et al. (2022), Ball et al. (2022), Bai et al. (2024), and Giannone and Primiceri (2024), but may also partly capture global demand dynamics as emphasized by Blanchard and Bernanke (2023). Meanwhile, the contributions from financial conditions and monetary policy likely represent the lagged effects of the exceptionally

accommodative stance maintained during the pandemic recovery, in line with Giannone and Primiceri (2024), who argue that early monetary accommodation temporarily reinforced inflationary momentum before the tightening cycle began.

The third phase covers the second half of 2022, when both models attribute pronounced downward pressures on inflation risk to monetary policy, consistent with the tightening cycle that began during this period.[15] On the inflationary side, both models now assign the largest contributions to inertial inflation, reflecting the carry-over from previously elevated price dynamics.

The final phase spans the remainder of the sample, covering forecast target vintages from 2023Q1 onward and encompassing the disinflation period. During this phase, both models attribute downward pressures on inflation risk to monetary policy, commodity prices, and the domestic business cycle, particularly in the QR model, broadly coinciding with the timing of receding demand shocks and the continuation of tight monetary policy documented in Giannone and Primiceri (2024). On the inflationary side, the dominant factor remains elevated past inflation, while inflation expectations also play a growing role, potentially reflecting the gradual adjustment of agents' beliefs following a period of persistently high inflation.[16]

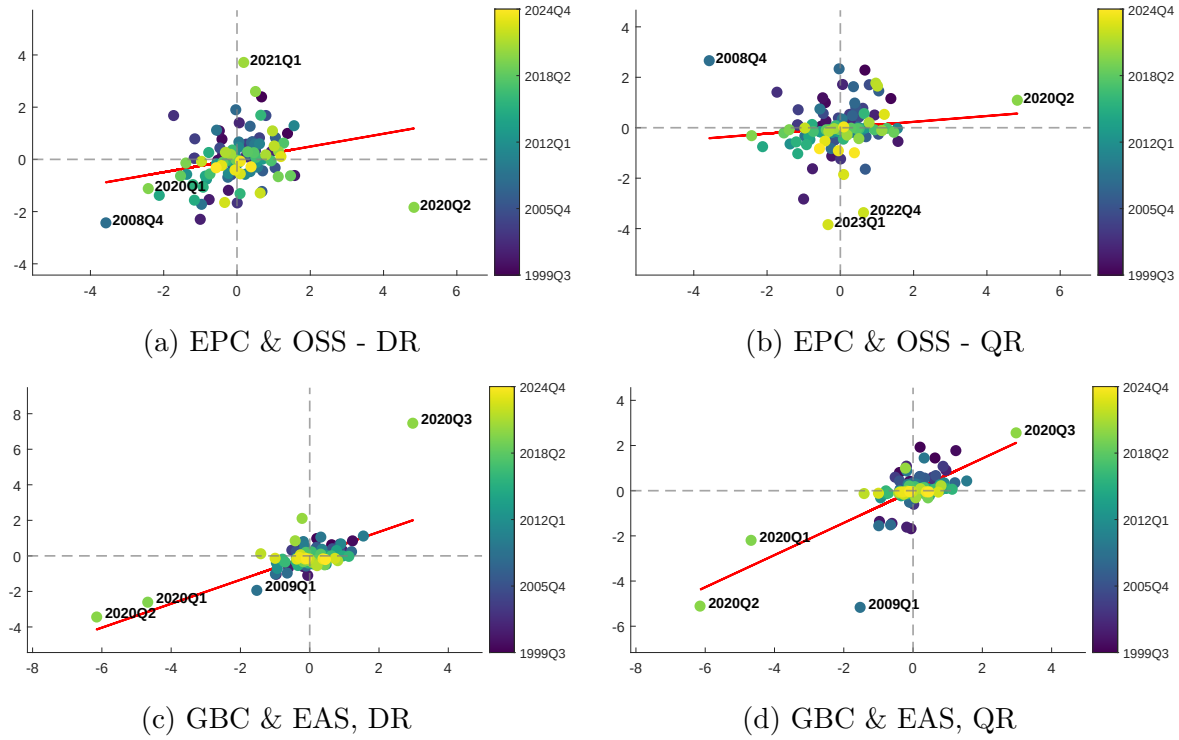
It is important to emphasize again that these decompositions are reduced-form in nature and should not be interpreted as evidence of causal relationships. Rather than identifying structural shocks, they serve to validate the proposed framework by assessing whether the model reacts plausibly to known macroeconomic developments. The close correspondence of the resulting patterns with causal and structural analyses in the post-pandemic inflation literature supports this interpretation. Moreover, these patterns remain largely robust across alternative inflation measures, as shown in the online appendix. Thus, while the exercise remains reduced-form, it captures the dynamic adjustment of the full predictive distribution in a preference-consistent manner and complements existing studies of realized inflation with a risk-based, real-time perspective on the post-pandemic inflation episode.

[15] While the decompositions appear to anticipate the official start of rate hikes, the analysis relies on model-based monetary policy indicators, which register tightening shocks irrespective of the level of interest rates.

[16] Section 5.4 in the online appendix repeats the exercise for $\alpha = \beta = 0$ as an additional validity check. The findings remain robust across models and preference parameter settings.

5.4 Additional Validation

Figure 6: Correlation with structural shocks for $h = 4$ and $\alpha = \beta = 2$



(a) EPC & OSS - DR

(b) EPC & OSS - QR

(c) GBC & EAS, DR

(d) GBC & EAS, QR

Notes: The top/bottom row compares the contribution of energy prices (EPC)/the global business cycle (GBC) from the risk decompositions to the structural oil supply shock (OSS)/economic activity shock (EAS) in Baumeister and Hamilton (2019), for $\alpha = \beta = 2$ in a scatter plot. The structural shocks are displayed on the x-axis. The red line indicates the regression line and the labelled points indicate the largest four outliers. The oil supply shock is inverted to account for the inverse relationship between oil prices and oil supply.

To further validate the decomposition results, the reduced-form Shapley attributions are compared with independently identified shocks from the structural macroeconomic literature. Figure 6 illustrates these relationships for two key channels. The top panel relates the contribution of energy prices to the oil supply shock of Baumeister and Hamilton (2019), while the bottom panel links the global business-cycle contribution to the corresponding economic activity shock from the same source. The red line shows the fitted regression, with the four largest outliers labelled for reference.

The contributions of energy prices are positively correlated with oil supply shocks, although the correlations are moderate, 0.25 for the DR model and 0.12 for the QR model. The main outliers correspond to the post-pandemic period and the global financial crisis. Excluding these observations increases the correlations to 0.32 and 0.23, respectively. For global economic activity, the relationships are considerably stronger, with correlations of 0.67 for the DR and 0.71 for the QR model. With outliers removed, the correlations

moderate to 0.37 and 0.43, but remain economically and statistically meaningful.

Despite being derived from purely empirical forecast densities, the decompositions thus mirror the behaviour of independently identified structural forces, reinforcing the credibility of the approach and its capacity to translate predictive distributions into economically interpretable contributions to inflation risk.

6 Conclusion

This paper develops a unified framework for measuring and decomposing macroeconomic risks from full predictive densities, integrating statistical modelling and policy-relevant risk evaluation within a single structure. The core contribution lies in the design of efficient, decomposition algorithms that attribute changes in both forecast densities and preference-weighted risk measures to their underlying economic predictors. These methods enhance interpretability while remaining fully consistent with the underlying predictive distribution and policy preferences.

A series of simulation exercises confirms the robustness of the framework across diverse data-generating processes, model structures, and sample sizes, demonstrating that the proposed algorithms yield accurate and stable decompositions even in complex settings. When applied to the recent U.S. inflation episode, the framework produces economically plausible assessments of the evolving balance of inflation risks, closely mirroring established structural and causal findings in the literature.

Because the approach is model-agnostic, it can be readily integrated with alternative density forecasting methods or extended to multivariate or cross-country applications. Future work may use the framework to analyse joint risk dynamics across macroeconomic aggregates, or to quantify policy trade-offs under multivariate uncertainty.

References

Aas, K., M. Jullum, and A. Løland (2021): "Explaining individual predictions when features are dependent: More accurate approximations to Shapley values," *Artificial Intelligence*, 298, 103502.

Abbai, B., O. Akinci, G. Benigno, J. di Giovanni, J. J. J. Groen, R. C. Heymann, L. Lin, and A. I. Noble (2022): "The Global Supply Side of Inflationary Pressures," Liberty Street Economics 20220128, Federal Reserve Bank of New York.

Adrian, T., N. Boyarchenko, and D. Giannone (2019): "Vulnerable Growth," *American Economic Review*, 109, 1263–1289.

Bai, X., J. Fernández-Villaverde, Y. Li, and F. Zanetti (2024): "The causal effects of global supply chain disruptions on macroeconomic outcomes: evidence and theory," Working Paper 32098, National Bureau of Economic Research.

Ball, L. M., D. Leigh, and P. Mishra (2022): "Understanding U.S. Inflation During the COVID Era," Working Paper 30613, National Bureau of Economic Research.

Banerjee, R., J. Contreras, A. Mehrotra, and F. Zampolli (2024): "Inflation at risk in advanced and emerging market economies," *Journal of International Money and Finance*, 103025.

Baumeister, C. and J. D. Hamilton (2019): "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks," *American Economic Review*, 109, 1873–1910.

Bekaert, G. and A. Popov (2019): "On the Link between the Volatility and Skewness of Growth," *IMF Economic Review*, 67, 746–790.

Bernanke, B. S. (2003): "Constrained Discretion and Monetary Policy," *BIS Review*, 5, 1–8.

Bhattacharyya, A., S. Pal, R. Mitra, and S. Rai (2022): "Applications of Bayesian Shrinkage Prior Models in Clinical Research with Categorical Responses," *BMC Medical Research Methodology*, 22, 126.

Blanchard, O. J. and B. S. Bernanke (2023): "What Caused the US Pandemic-Era Inflation?" Working Paper 31417, National Bureau of Economic Research.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (2017): "Variational Inference: A Review for Statisticians," *Journal of the American Statistical Association*, 112, 859–877.

Blinder, A. S. (1997): "Distinguished Lecture on Economics in Government: What Central Bankers Could Learn from Academics–And Vice Versa," *Journal of Economic Perspectives*, 11, 3–19.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010): "The Horseshoe Estimator for Sparse Signals," *Biometrika*, 97, 465–480.

Cascaldi-Garcia, D., L. Guerrieri, M. Iacoviello, and M. Modugno (2024): "Lessons from the Co-movement of Inflation around the World," Feds notes, Board of Governors of the Federal Reserve System.

Chen, H., J. D. Janizek, S. Lundberg, and S.-I. Lee (2020): "True to the model or true to the data?" *arXiv preprint arXiv:2006.16234*.

Ding, D. and G. Karabatsos (2021): "Dirichlet process mixture models with shrinkage prior," *Stat*, 10, e371.

Frühwirth-Schnatter, S. (2006): *Finite Mixture and Markov Switching Models*, Springer Series in Statistics, New York, NY: Springer.

Giannone, D. and G. Primiceri (2024): "The drivers of post-pandemic inflation," Working Paper 32859, National Bureau of Economic Research.

Gilchrist, S. and E. Zakrajšek (2012): "Credit Spreads and Business Cycle Fluctuations," *American Economic Review*, 102, 1692–1720.

Goodfriend, M. (1993): "Interest rate policy and the inflation scare problem: 1979-1992," *FRB Richmond Economic Quarterly*, 79, 1–23.

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

Ha, J., M. A. Kose, and F. Ohnsorge (2021): "Inflation During the Pandemic: What Happened? What Is Next?" Discussion Paper 3886814, CEPR.

Hall, S. G., G. S. Tavlas, and Y. Wang (2023): "Drivers and Spillover Effects of Inflation: The United States, the Euro Area, and the United Kingdom," *Journal of International Money and Finance*, 131, 102776.

Janzing, D., L. Minorics, and P. Blöbaum (2020): "Feature relevance quantification in explainable AI: A causal problem," in *International Conference on artificial intelligence and statistics*, PMLR, 2907–2916.

Kilian, L. and S. Manganelli (2007): "Quantifying the Risk of Deflation," *Journal of Money, Credit and Banking*, 39, 561–590.

——— (2008): "The Central Banker as a Risk Manager: Estimating the Federal Reserve's Preferences under Greenspan," *Journal of Money, Credit and Banking*, 40, 1103–1129.

Koenker, R. (2005): *Quantile regression*, vol. 38, Cambridge university press.

Koenker, R. and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33.

Koop, G. and D. Korobilis (2023): "Bayesian Dynamic Variable Selection in High Dimensions," *International Economic Review*, 64, 1047–1074.

Koop, G. M. (2013): "Forecasting with Medium and Large Bayesian VARS," *Journal of Applied Econometrics*, 28, 177–203.

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

Korobilis, D., B. Landau, A. Musso, and A. Phella (2021): "The Time-Varying Evolution of Inflation Risks," Working Paper Series 2600, European Central Bank.

Korobilis, D. and M. Schröder (2025a): "Monitoring multi-country macroeconomic risk: A quantile factor-augmented vector autoregressive (QFAVAR) approach," *Journal of Econometrics*, 249, 105730.

——— (2025b): "Probabilistic Quantile Factor Analysis," *Journal of Business & Economic Statistics*, 43, 530–543.

Korobilis, D. and K. Shimizu (2022): "Bayesian Approaches to Shrinkage and Sparse Estimation," *Foundations and Trends® in Econometrics*, 11, 230–354.

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

LaBelle, J. and A. M. Santacreu (2022): "Global Supply Chain Disruptions and Inflation during the COVID-19 Pandemic," *Federal Reserve Bank of St. Louis Review*, 104(2), 78–91.

Labonne, P. (2025): "Asymmetric uncertainty: Nowcasting using skewness in real-time data," *International Journal of Forecasting*, 41, 229–250.

Lim, D., B. Park, D. Nott, X. Wang, and T. Choi (2020): "Sparse Signal Shrinkage and Outlier Detection in High-Dimensional Quantile Regression with Variational Bayes," *Statistics and Its Interface*, 13, 237–249.

Lopez-Salido, D. and F. Loria (2024): "Inflation at risk," *Journal of Monetary Economics*, 103570.

Machina, MJ. and M. Rothschild (1987): "Risk," *The New Palgrave Dictionary of Economics*.

Makalic, E. and D. F. Schmidt (2016): "A Simple Sampler for the Horseshoe Estimator," *IEEE Signal Processing Letters*, 23, 179–182.

Malsiner-Walli, G., S. Frühwirth-Schnatter, and B. Grün (2016): "Model-based clustering based on sparse finite Gaussian mixtures," *Statistics and computing*, 26, 303–324.

Manzan, S. (2015): "Forecasting the Distribution of Economic Variables in a Data-Rich Environment," *Journal of Business & Economic Statistics*, 33, 144–164.

McCracken, M. W. and S. Ng (2020): "FRED-QD: A Quarterly Database for Macroeconomic Research," Working Paper 26872, National Bureau of Economic Research.

Miranda-Agrippino, S. (2016): "Unsurprising Shocks: Information, Premia, and the Monetary Transmission," Staff Working Paper 626, Bank of England.

Mitchell, J., A. Poon, and D. Zhu (2024): "Constructing density forecasts from quantile regressions: Multimodality in macrofinancial dynamics," *Journal of Applied Econometrics*, 39, 790–812.

Monache, D. D., A. D. Polis, and I. Petrella (2024): "Modeling and Forecasting Macroeconomic Downside Risk," *Journal of Business & Economic Statistics*, 42, 1010–1025.

Ormerod, J. T., C. You, and S. Müller (2017): "A Variational Bayes Approach to Variable Selection," *Electronic Journal of Statistics*, 11, 3549–3594.

Patton, A. J., J. F. Ziegel, and R. Chen (2019): "Dynamic semiparametric models for expected shortfall (and Value-at-Risk)," *Journal of Econometrics*, 211, 388–413.

Polson, N. G., J. G. Scott, and J. Windle (2013): "Bayesian Inference for Logistic Models Using Polya-Gamma Latent Variables," *Journal of the American Statistical Association*, 108, 1339–1349.

Reichlin, L., G. Ricco, and T. Hasenzagl (2020): "Financial Variables as Predictors of Real Growth Vulnerability," Deutsche Bundesbank Discussion Paper 05/2020, Deutsche Bundesbank.

Rigon, T. and D. Durante (2021): "Tractable Bayesian density regression via logit stick-breaking priors," *Journal of Statistical Planning and Inference*, 211, 131–142.

Rossi, B. and T. Sekhposyan (2019): "Alternative tests for correct specification of conditional predictive densities," *Journal of Econometrics*, 208, 638–657.

Shapley, L. S. (1953): *A Value for n-Person Games, Contributions to the Theory of Games (AM-28), Volume II*, Princeton: Princeton University Press, 307–317.

Stock, J. H. and M. W. Watson (2007): "Why has US inflation become harder to forecast?" *Journal of Money, Credit and Banking*, 39, 3–33.

——— (2012): "Disentangling the Channels of the 2007-2009 Recession," Working Paper 18094, National Bureau of Economic Research.

Strumbelj, E. and I. Kononenko (2010): "An Efficient Explanation of Individual Classifications Using Game Theory," *The Journal of Machine Learning Research*, 11, 1–18.

Sundararajan, M., A. Taly, and Q. Yan (2017): "Axiomatic attribution for deep networks," in *International conference on machine learning*, PMLR, 3319–3328.

Svensson, L. E. O. (2002): "Inflation Targeting: Should It Be Modeled as an Instrument Rule or a Targeting Rule?" *European Economic Review*, 46, 771–780.

Tagliabracci, A. (2020): "Asymmetry in the Conditional Distribution of Euro-area Inflation," Working Paper 1270, Bank of Italy.

Villani, M., R. Kohn, and P. Giordani (2009): "Regression Density Estimation Using Smooth Adaptive Gaussian Mixtures," *Journal of Econometrics*, 153, 155–173.

Wei, R. and S. Ghosal (2020): "Contraction Properties of Shrinkage Priors in Logistic Regression," *Journal of Statistical Planning and Inference*, 207, 215–229.

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

Online Appendix to "Mixing it up: Inflation at risk"

Maximilian Schröder

1 Proof of Lemma 3.1

Proof: Let $f_x(y)_{t+h|t}$ denote the forecast distribution of a variable at time t for horizon $t+h$. Further, let $y^\star$ denote a set of g ordered grid points across the entire support of y. $y_i^\star$ and $y_{i-1}^\star$ hence denote two neighbouring grid points such that $y_{i-1}^\star < y_i^\star$. Further, let $y_1^\star = -\infty$ and $y_g^\star = \infty$. Then

$$\int_{-\infty}^{\infty} f_x(y)_{t+h|t} dy = \sum_{i=2}^{g} \int_{y_{i-1}^\star}^{y_i^\star} f_x(y)_{t+h|t} dy = 1, \text{ and } \int_{-\infty}^{\infty} f_x(y)_{t+h} dy = \sum_{i=2}^{g} \int_{y_{i-1}^\star}^{y_i^\star} f_x(y)_{t+h} dy = 1.$$

Given the definition of the predictor contributions

$$\sum_{j}^{n} \varphi_j^{y_i^\star} + \int_{y_{i-1}^\star}^{y_i^\star} f_x(y)_{t+h} dy = \int_{y_{i-1}^\star}^{y_i^\star} f_x(y)_{t+h|t} dy.$$

As a result

$$\int_{-\infty}^{\infty} f_x(y)_{t+h|t} dy = \sum_{i=2}^{g} \int_{y_{i-1}^\star}^{y_i^\star} f_x(y)_{t+h|t} dy = \sum_{i=2}^{g} \left[\sum_{j}^{n} \varphi_j^{y_i^\star} + \int_{y_{i-1}^\star}^{y_i^\star} f_x(y)_{t+h} dy \right] = 1. \quad \blacksquare$$

2 Estimation Algorithms

2.1 Variational Bayes Inference for Density Regression

Variational Bayes (VB) and MCMC both provide approximations to a given posterior. However; while MCMC relies on sampling, VB approximates the posterior through solving an optimization problem. For a general introduction to VB see Blei et al. (2017). Because this optimization problem requires fewer iterations than sampling from an MCMC chain, VB's main advantage lies in its computational speed. As a caveat, whereas MCMC guarantees exact draws from the posterior, VB yields a solution close to the posterior. As such, VB is particularly useful for tasks where computational time is critical and

accurate parameter estimates are less of a concern, such as, big-data problems, models with highly intractable posteriors, real-time monitoring tasks, or forecasting exercises. The algorithm proposed in this paper extends the VB algorithm in Rigon and Durante (2021) by introducing shrinkage into the sequential logistic regressions and evolves as follows:

1. Update the variational density of the mixture component assignment indicators $\zeta_{t,c}$. Here the model is augmented using the binary indicators z instead of the mixture component membership indicators G_t.
2. Update the variational densities corresponding to the regression coefficients in the logistic regressions, $\boldsymbol{\psi}_c$, for all mixture components.
3. Update the variational density of the horseshoe prior coefficients.
4. Update the variational density of the Pólya-Gamma indicators $\omega_{t,c}$.
5. Update the variational density of the regression coefficients, $\boldsymbol{\beta}_c$, and the precision parameters, τ_c, of the individual mixture components.

Throughout lower case letters denote scalars, bold lower case letters denote vectors and bold capital letters denote matrices. To ease notation, whenever indices referring to the mixture assignment indicators, time periods, or variables are suppressed, they are collected in a single vector. Collect all model parameters in $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{z}, \boldsymbol{\psi}, \boldsymbol{\omega}, \boldsymbol{\nu}, \bar{\boldsymbol{\tau}}^2, \boldsymbol{\eta}, \boldsymbol{\xi})$. For a family of tractable densities $q(\boldsymbol{\Theta})$, we aim to find a density $q^\star$ that best approximates the posterior $p(\boldsymbol{\Theta}|\boldsymbol{x})$ by minimizing

$$q_x^\star(\boldsymbol{\Theta}) = \underset{q \in \mathcal{Q}}{\operatorname{argmin}} \, \mathbb{D}_{KL} \left(q(\boldsymbol{\Theta}) || p(\boldsymbol{\Theta}|\boldsymbol{x}) \right). \tag{1}$$

This is equivalent to maximizing

$$ELBO = \mathbb{E}_{q_x(\boldsymbol{\Theta})} \left[\log p(\boldsymbol{x}) \right] + \mathbb{E}_{q_x(\boldsymbol{\Theta})} \left[\log p(\boldsymbol{\Theta}) \right] - \mathbb{E}_{q_x(\boldsymbol{\Theta})} \left[\log q_x(\boldsymbol{\Theta}) \right], \tag{2}$$

where KL denotes the Kullback-Leibler divergence. Importantly, the solution requires optimizing over a family of distribution functions and hence the application of variational calculus. A key part of VB is to simplify the variational posterior by factorization, such that the resulting optimization problem becomes tractable. On the one hand, the simpler the resulting posterior, the easier the optimization problem. On the other hand, this implies independence assumptions between the individual variational components and hence leads to approximation error. This gives rise to a trade-off (Ormerod et al., 2017). In the proposed estimator, the so-called mean-field factorization is applied. This yields

$$q_x(\Theta) = \prod_{c=1}^{C-1} q_x(\boldsymbol{\psi}_c) \prod_{c=1}^{C-1} q_x(\xi_c) \prod_{c=1}^{C-1} \prod_{j=1}^{n} q_x(\nu^2_{j,c}) \prod_{c=1}^{C-1} \prod_{j=1}^{n} q_x(\bar{\tau}^2_{j,c}) \prod_{c=1}^{C-1} \prod_{j=1}^{n} q_x(\eta^2_{j,c})$$
$$\cdot \prod_{c=1}^{C} q_x(\boldsymbol{\beta}_c) \prod_{c=1}^{C} q_x(\tau_c) \prod_{c=1}^{C-1} \prod_{t=1}^{T} q_{x_t}(z_{t,c}) \prod_{c=1}^{C-1} \prod_{t=1}^{T} q_{x_t}(\omega_{t,c}). \tag{3}$$

This implies independence between the regression coefficients in the logistic regressions and kernel regressions, the cluster assignment indicators, the parameters of the horseshoe prior, and the Pólya-Gamma indicators. Additionally, conditional independence between the priors is assumed.

Given these assumptions, the solution to the optimization problem can be obtained by sequentially iterating over the densities

$$q_x(\boldsymbol{\Theta}_l) \propto \exp\ \mathbb{E}_{q_x(\boldsymbol{\Theta}_{(-l)})} \left(\log\ p(\boldsymbol{\Theta}_l | \boldsymbol{\Theta}_{(-l)}, \boldsymbol{x})\right) \tag{4}$$

where $\boldsymbol{\Theta}_{(-l)}$ denotes all elements of $\boldsymbol{\Theta}$, excluding those in the l^{th} group, $l = 1, \cdots, L$.

As a result, the variational posterior can be obtained by calculating the variational expectation of the conditional posterior. As described above, the accuracy of the variational approximation hinges on how well the partitioning matches the independence structure of the parameters in the target posterior. For a general discussion of this issue see e.g. Ormerod et al. (2017). To arrive at the final variational densities, insert 3 and the priors into 4. The final algorithm is given in algorithm 1.

Algorithm 1: Variational Bayes algorithm

[0] Initialize the model parameters. Let $q^m(.)$ denote a generic variational distribution at iteration m.

begin

[1] Compute the variational density $q^*_{\boldsymbol{x}_t}(z_{i,c})$ for each $t = 1, \dots, T$ and $c = 1, \dots, C-1$:

for $t = 1 : T$ **do**

for $c = 1 : C$ **do**

Following Rigon and Durante (2021), $q^*_{x_t}(z_{i,h})$ coincides with the probability mass function of $Bern(\rho_{i,h})$, where

$$\rho_{i,h} = \frac{1}{1 + exp\left(-\left[\boldsymbol{x}_t'\mathbb{E}(\phi_c) + \sum_{l=c}^{C} \zeta_{t,l}^c \left(0.5 \cdot \mathbb{E}(\log \tau_l) - 0.5 \cdot \mathbb{E}(\log \tau_l) \cdot \mathbb{E}(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_l)^2\right)\right]\right)}.$$

with $\zeta_{t,l}^c = \prod_{r=1}^{l-1}(1-\rho_{t,r})$ for $l = c$ and $\zeta_{t,l}^c = -\rho_{t,l}\prod_{r=1, r\neq c}^{l-1}(1-\rho_{t,r})$, and $\rho_{t,C} = 1$.

Assign: $\mathbb{E}(\zeta_{t,c}) = \mathbb{E}\left(z_{t,c}\prod_{l=1}^{c-1}(1-z_{t,l})\right) = \rho_{t,c}\prod_{l=1}^{c-1}(1-\rho_{t,l})$, and $\mathbb{E}(z_{t,c}) = \rho_{t,c}$.

end

end

[2] Update $q^*_x(\boldsymbol{\psi}_c)$ and $q^*_x(\boldsymbol{\upsilon}_c^2)$, $q^*_x(\bar{\boldsymbol{\tau}}_c^2)$, $q^*_x(\boldsymbol{\eta}_c)$, and $q^*_x(\xi_c)$ for $c = 1, \dots, C-1$:

for $c = 1, \dots, C-1$ **do**

[2a] Update $q^*_x(\boldsymbol{\psi}_c)$ from $N(\boldsymbol{\mu}_{\psi,c}, \boldsymbol{\Sigma}_{\psi,c})$, with
$\boldsymbol{\Sigma}_{\psi,c} = \left(\boldsymbol{X}'\boldsymbol{V}_c\boldsymbol{X} + \boldsymbol{\Lambda}^{-1}\right)^{-1}$, $\boldsymbol{\mu}_{\psi,c} = \boldsymbol{\Sigma}_{\psi,c}\left(\boldsymbol{X}'\boldsymbol{\kappa}_c\right)$, with $\boldsymbol{\kappa}_c = (\mathbb{E}(z_{1,c}) - 0.5, \dots, \mathbb{E}(z_{T,c}) - 0.5)'$,
$\boldsymbol{V}_c = diag\left(\mathbb{E}(\omega_{1,c}), \dots, \mathbb{E}(\omega_{T,c})\right)$, and $\boldsymbol{\Lambda}_c^{-1} = diag\left(\mathbb{E}(\boldsymbol{\upsilon}_c^{-2})\mathbb{E}(\boldsymbol{\tau}_c^{-2})\right)$.
Assign: $\mathbb{E}(\boldsymbol{\psi}_c) = \boldsymbol{\mu}_{\psi,c}$ and $\mathbb{E}(\boldsymbol{\psi}_c^2) = \boldsymbol{\mu}_{\psi,c}^2 + diag(\boldsymbol{\Sigma}_{\psi,c})$.

[2b] Update $q^*_x(\boldsymbol{\upsilon}_c^2)$, $q^*_x(\bar{\boldsymbol{\tau}}_c^2)$, $q^*_x(\boldsymbol{\eta}_c)$, and $q^*_x(\xi_c)$ for $c = 1, \dots, C-1$:
$q^*(\upsilon_{j,c}^2) = IG(a_{\upsilon,c}, b_{\upsilon,c}) = IG\left(1, \frac{\mathbb{E}(\psi_{j,c}^2)}{2} + \mathbb{E}\left(\frac{1}{\eta_{j,c}}\right)\right)$,
$q^*(\eta_{j,c}) = IG(a_{\eta,c}, b_{\eta,c}) = IG\left(1, \mathbb{E}\left(\frac{1}{\upsilon_{j,c}^2}\right) + b_\psi^{-2}\mathbb{E}\left(\frac{1}{\bar{\tau}_{j,c}^2}\right)\right)$,
$q^*(\bar{\tau}_{j,c}^2) = IG(a_{\bar{\tau},c}, b_{\bar{\tau},c}) = IG\left(1, b_\psi^{-2}\mathbb{E}\left(\frac{1}{\eta_{j,c}}\right) + \mathbb{E}\left(\frac{1}{\xi_c}\right)\right)$,
$q^*(\xi_c) = IG(a_{\xi,c}, b_{\xi,c}) = IG\left(\frac{n+1}{2}, 1 + \sum_{j=1}^{n}\mathbb{E}\left(\frac{1}{\bar{\tau}_{j,c}^2}\right)\right)$,
for $j = 1, \cdots, n$.
Assign: $\mathbb{E}(\upsilon_{j,c}^2) = \frac{a_{\upsilon,c}}{b_{\upsilon,c}}$, $\mathbb{E}(\eta_{j,c}) = \frac{a_{\eta,c}}{b_{\eta,c}}$, $\mathbb{E}(\bar{\tau}_{j,c}) = \frac{a_{\tau,c}}{b_{\tau,c}}$, and $\mathbb{E}(\xi_c) = \frac{a_{\xi,c}}{b_{\xi,c}}$.

end

[3] Update $q^*_x(\omega_{t,c})$ for $t = 1, \dots, T$ and $c = 1, \dots, C$.

for $t = 1 : T$ **do**

for $c = 1 : C$ **do**

Update $q^*_x(\omega_{t,c}) \sim PG(1, \delta_{t,c})$, with $\delta_{t,c}^2 = \boldsymbol{x}_t'\mathbb{E}(\boldsymbol{\psi}_c^2)\boldsymbol{x}_t$.
Assign: $\mathbb{E}(\omega_{t,c}) = 0.5 \cdot \delta_{t,c}^{-1} tanh(0.5 \cdot \delta_{t,c})$.

end

end

[4] Update $q^*_x(\boldsymbol{\beta}_c)$ and $q^*_x(\tau_c)$ for $c = 1, \dots, C$:

for $c = 1 : C$ **do**

Update $q^*_x(\boldsymbol{\beta}_c) \sim N(\boldsymbol{\mu}_{\beta_c}, \boldsymbol{\Sigma}_{\beta_c})$ and $q^*_x(\tau_c) \sim IG(a_{\tau_c}, b_{\tau_c})$, with $\boldsymbol{\Sigma}_{\beta_c} = \left(\boldsymbol{X}'\boldsymbol{\Gamma}_c\boldsymbol{X} + \underline{\boldsymbol{\Sigma}}_\beta^{-1}\right)^{-1}$,
$\boldsymbol{\mu}_{\beta_c} = \boldsymbol{\Sigma}_{\beta_c}\left(\boldsymbol{X}\boldsymbol{\Gamma}_c\boldsymbol{y} + \underline{\boldsymbol{\Sigma}}_\beta^{-1}\underline{\boldsymbol{\mu}}_\beta\right)$, $a_{\tau_c} = a_\tau + 0.5\sum_{t=1}^{T}\mathbb{E}(\zeta_{t,c})$, and
$b_{\tau_c} = b_\tau + 0.5\sum_{t=1}^{T}\mathbb{E}(\zeta_{t,c})\mathbb{E}\left((y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_c)^2\right)$, with $\Gamma_c = \mathbb{E}(\tau_c)diag(\mathbb{E}(\zeta_{1,c}), \dots, \mathbb{E}(\zeta_{T,c}))$.
Assign: $\mathbb{E}(\boldsymbol{\beta}_c) = \boldsymbol{\mu}_{\beta_c}$, $\mathbb{E}(\boldsymbol{\beta}_c^2) = \boldsymbol{\mu}_{\beta_c}^2 + diag\left(\boldsymbol{\Sigma}_{\beta_c}\right)$, and $\mathbb{E}(\tau_c) = \frac{a_{\tau_c}}{b_{\tau_c}}$.

end

end

Finally, note that a different parametrization of the horseshoe prior is used:

$$\begin{aligned}
\boldsymbol{\psi}_c | \left\{\nu_{j,c}^2, \eta_{j,c}, \bar{\tau}_{j,c}^2\right\}_{j=1}^n, \xi_c &\sim N(\mathbf{0}, \boldsymbol{\Lambda}_c), \\
\bar{\lambda}_{j,c}^2 | \upsilon_{j,c} &\sim G^{-1}\left(\frac{1}{2}, \frac{1}{\upsilon_{j,c}}\right), \quad \text{for } j = 1, \cdots, n, \\
\upsilon_{j,c} | \bar{\tau}_{j,c}^2 &\sim G^{-1}\left(\frac{1}{2}, \frac{1}{b_\psi^2 \tau_{j,c}^2}\right), \quad \text{for } j = 1, \cdots, n, \\
\bar{\tau}_{j,c}^2 | \xi_c &\sim G^{-1}\left(\frac{1}{2}, \frac{1}{\xi_c}\right), \\
\xi_c &\sim G^{-1}\left(\frac{1}{2}, 1\right),
\end{aligned} \tag{5}$$

where b_ψ is a hyperparameter.

2.2 Implementation Details - Quantile Regression and UC-SV

This appendix provides details on the estimation of the benchmark models used in the forecasting exercise in the main body of the paper. All benchmarks can be obtained as special cases of the Variational Bayes algorithm for the density regression model in algorithm 1.

QR: The QR model can be estimated as a special case of step [4] in algorithm 1, by setting $\mathbb{E}(\zeta_{tc}), \ldots, \mathbb{E}(\zeta_{Tc}) = 1$. The horseshoe prior can be imposed by modifying step [2b] of algorithm 1 , by setting $\psi_c = \beta_c$, and by replacing $\underline{\Sigma}_\beta^{-1}$ in step [4] by Λ^{-1} from step [2a]. In the QR model the residuals have an Asymmetric Laplace distributions. However, following Lim et al. (2020) and others the Asymmetric Laplace distribution can be rewritten as a normal-exponential mixture. For quantile level q, define $\kappa_1 = (1-2q)/(q(1-q))$ and $\kappa_2 = 2/(q(1-q))$. The modified regression equation reads $y_t = \boldsymbol{\beta}(q)'\boldsymbol{x}_t + \kappa_1 z_t(q) + \kappa_2\sqrt{\sigma(q) z_t(q)}\varepsilon_t$. In step [4] of algorithm 1, $\Gamma_c = 1/\kappa_2^2 \mathbb{E}(1/\sigma) diag(\mathbb{E}(1/z_1(q)), \ldots, \mathbb{E}(1/z_T(q)))$. In addition, $\mu(q)_{\beta_c} = \Sigma_\beta(q)[\boldsymbol{x}'\Gamma_c \boldsymbol{y} - \kappa_1/\kappa_2^2 \mathbb{E}(1/\sigma(q)) \sum_{t=1}^T \boldsymbol{x}_t]$. z_t is given by the inverse Gaussian distribution with $z_t(q) \sim InG(0.5, a^z(q), b_t^z(q))$, where $a^z(q) = \mathbb{E}(1/\sigma(q))(2 + \kappa_1/\kappa_2^2)$, $b_t^z(q) = \mathbb{E}(1/\sigma(q))[1/\kappa_2^2((\boldsymbol{y}_t - \boldsymbol{x}_t\beta(q))^2 + \boldsymbol{x}_t diag(\Sigma_\beta(q))\boldsymbol{x}_t')]$. The expectations $\mathbb{E}(z_t) = (\sqrt{b_t^z(q)}K_{3/2}(\sqrt{a_t^z(q)b_t^z(q)}))/(\sqrt{a_t^z(q)}K_{1/2}(\sqrt{a_t^z(q)b_t^z(q)}))$ and $\mathbb{E}(1/z_t) = (\sqrt{a_t^z(q)}K_{3/2}(\sqrt{a_t^z(q)b_t^z(q)}))/(\sqrt{b_t^z(q)}K_{1/2}(\sqrt{a_t^z(q)b_t^z(q)})) - 1/b_t^z(q)$, where $K_p(\bullet)$ is the Bessel function of order p. Finally, for $\sigma \sim IG(a_\sigma(q), b_\sigma(q))$ the

updates are $a_\sigma(q) = \bar{a}_\sigma(q) + 3T$ and $b_\sigma(b) = \bar{b}_\sigma(q) + \sum_{t=1}^{T}[\mathbb{E}(1/z_t(q))((y_t - \mathbb{E}(\boldsymbol{\beta}(q))'\boldsymbol{x}_t)^2 + \boldsymbol{x}_t\Sigma_\beta(q)\boldsymbol{x}_t')/(2\kappa_2^2) - \kappa_1(y_t - \mathbb{E}(\boldsymbol{\beta}(q))'\boldsymbol{x}_t)/(\kappa_2^2) + (1 + \kappa_1^2/2\kappa_2^2)\mathbb{E}(z_t(q))]$, where $\bar{a}_\sigma(q)$ and $\bar{b}_\sigma(q)$ are the priors. Expectations are computed analogously to the other cases.

UC-SV: To implement the UC-SV model as a special case of step [4] in algorithm 1, it first has to be rewritten as a TVP-AR model, with a time-varying constant as the only exogenous variable. This TVP regression can then be rewritten as a static regression model with $T \cdot 1$ parameters as described in Korobilis and Shimizu (2022). In particular, the regression model is then given by

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1 & 0 & \dots & \dots & 0 \\ \boldsymbol{x}_2 & \boldsymbol{x}_2 & \ddots & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{x}_T & \boldsymbol{x}_T & \dots & \dots & \boldsymbol{x}_T \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \Delta\boldsymbol{\beta}_2 \\ \vdots \\ \Delta\boldsymbol{\beta}_T \end{bmatrix} + \varepsilon_t,$$

where $\boldsymbol{x}_t = 1$ in case of the UC-SV. This overparametrized regression can then be computed by imposing shrinkage priors such as the horseshoe prior on the coefficients by modifying step [2b] of algorithm 1 and setting $\psi_c = \beta_c$, and by replacing $\underline{\Sigma}_\beta^{-1}$ in step [4] by Λ^{-1} from step [2a]. Estimates of $\boldsymbol{\beta}_t$ are obtained by computing the cumulative sum of the parameter vector across time for all variables. In addition, in step [4] in algorithm 1, $\mathbb{E}(\zeta_{tc}), \dots, \mathbb{E}(\zeta_{Tc})$ needs to be set to 1. To incorporate stochastic volatility, the algorithm follows Koop and Korobilis (2023). In particular, the variational expectation of time-varying precision parameter $\tau_t \sim Gamma(\delta a_{t-1}, \delta a_{t-1})$, is $\mathbb{E}(\tau_t) = a_t/b_t$, where $a_t = 0.5 + \delta a_{t-1}$, $b_t = 0.5 \cdot \varepsilon_t^2 + \delta b_{t-1}$, and δ is a smoothing parameter set to 0.8, following Koop and Korobilis (2023), and ε_t is the regression residual. Smoothed precision parameters, τ_t^s, can be obtained by iteratively computing $\tau_t^s = (1 - \delta)\tau_t + \delta\tau_{t+1}$ backwards through time.

3 Decomposition Algorithms

This section gives the full algorithms for computing the contributions introduced in the main body of the text. Specifically, algorithm 2 allows to decompose the forecast density, while algorithm 3 decomposes the risk measures.

For computational efficiency, the replication codes implement the above algorithms

Algorithm 2: Forecast Density Decompositions

[0] Specify the number of samples M, and the number of grid points, g. Generate a grid for $\hat{y}$, with $\hat{y}_i \in [\lfloor y_{1:T} \rfloor, ..., \lceil y_{1:T} \rceil]$ for $i = 1, \ldots, g$.
[1] Compute the contribution for variable j to the forecast distribution for $t+h$ at time t.
begin
  [a] Iterate over random samples. The sampling population is the set of all time-period/variable permutation pairs.
  for $m = 1 : M$ **do**
    [i]Sample a random time period, $t^* \in_R [1 : t-1]$, from the the preceding sample of observations and set $\boldsymbol{z}_t = \boldsymbol{x}_{t^*}$.
    [ii] Sample a random permutation P of $\{1, \ldots, n\}$. Define the coalition $S\{k \in \{1, \ldots, n\} \setminus \{j\} : P(k) < P(j)\}$, i.e., the set of features that precede j in P.
    [iii] Create two new data vectors. For the first vector, replace the values in $\boldsymbol{z}_t$ with the values in $\boldsymbol{x}_t$ for the indices in S. For the second vector, also replace $z_{t,j}$ with $x_{t,j}$. As an example, let $S = [2, \ldots, j-1]$. This yields:

$$\boldsymbol{x}_t^{-j} = [z_{t,1}, x_{t,2}, \ldots, x_{t,j-1}, z_{t,j}, z_{t,j+1}, \ldots, z_{t,n}]$$
$$\boldsymbol{x}_t^{+j} = [z_{t,1}, x_{t,2}, \ldots, x_{t,j-1}, x_{t,j}, z_{t,j+1}, \ldots, z_{t,n}].$$

    [iv] Compute the contribution for j between the grid points $\hat{\pi}_i$ and $\hat{\pi}_{i-1}$ at draw m

$$V_{m,j}^i = \int_{\hat{y}_{i-1}}^{\hat{y}_i} f_x(y)_{t+h|t}^{+j} dy - \int_{\hat{y}_{i-1}}^{\hat{y}_i} f_x(y)_{t+h|t}^{-j} dy \qquad \text{for every} \qquad i = 2, \ldots, g.$$

  end
  [b] Compute the contribution of j between grid points $\hat{y}_i$ and $\hat{y}_{i-1}$, $\hat{\varphi}_j^i$, by taking the average over all samples, M.

$$\hat{\varphi}_j^i = \frac{1}{M} \sum_m^M V_{m,j}^i.$$

end
Notes: The integrals can be computed with quadrature methods.

slightly differently. Instead of proceeding by variable, each iteration, m, a random time period and random order of variables is sampled: e.g. $S = [5, 1, 2, 4, 3]$, for $n = 5$. The inner loop then replaces the observations recursively for $j = 1 : n$ - following this random order - and records the marginal contributions accordingly, i.e. for variable $S(j)$ at iteration j. Because the $\boldsymbol{x}_t^{-j}$ is not reset between iterations of the inner loop, this is equivalent to sampling random permutations.

In addition, the replication files allow to compute attributions for variable groups instead of individual variables and avoid computing attributions for constants to reduce the computational burden and facilitate the application to high dimensional problems.

Algorithm 3: Risk Measure Decompositions

[0] Specify the number of samples M, and set α, β, $\bar{\pi}$, $\underline{\pi}$.
[1] Compute the contribution for variable j to the risk measures for $t+h$ at time t.
begin
  [a] Iterate over random samples. The sampling population is the set of all time-period/variable permutation pairs.
  for $m = 1 : M$ **do**
    [i]Sample a random time period, $t^* \in_R [1 : t-1]$, from the the preceding sample of observations and set $\boldsymbol{z}_t = \boldsymbol{x}_{t^*}$.
    [ii] Sample a random permutation P of $\{1, \ldots, n\}$. Define the coalition $S\{k \in \{1, \ldots, n\} \setminus \{j\} : P(k) < P(j)\}$, i.e., the set of features that precede j in P.
    [iii] Create two new data vectors. For the first vector, replace the values in $\boldsymbol{z}_t$ with the values in $\boldsymbol{x}_t$ for the indices in S. For the second vector, also replace $z_{t,j}$ with $x_{t,j}$. As an example, let $S = [2, \ldots, j-1]$. This yields:

$$\boldsymbol{x}_t^{-j} = [z_{t,1}, x_{t,2}, \ldots, x_{t,j-1}, z_{t,j}, z_{t,j+1}, \ldots, z_{t,n}]$$
$$\boldsymbol{x}_t^{+j} = [z_{t,1}, x_{t,2}, \ldots, x_{t,j-1}, x_{t,j}, z_{t,j+1}, \ldots, z_{t,n}].$$

    [iv] Compute the contribution for j to DR_α and EIR_β at draw m.

$$V_{m,j}^{DR_\alpha} = -\left(\int_{-\infty}^{\bar{\pi}} (\underline{\pi} - \pi)^\alpha f_x(\pi)_{t+h|t}^{+j} d\pi - \int_{-\infty}^{\bar{\pi}} (\pi - \bar{\pi})^\alpha f_x(\pi)_{t+h|t}^{-j} d\pi\right)$$
$$V_{m,j}^{EIR_\beta} = \int_{\bar{\pi}}^{\infty} (\underline{\pi} - \pi)^\beta f_x(\pi)_{t+h|t}^{+j} d\pi - \int_{\bar{\pi}}^{\infty} (\pi - \bar{\pi})^\beta f_x(\pi)_{t+h|t}^{-j} d\pi$$

  end
  [b] Compute the contribution of j to DR_α and EIR_β, $\hat{\varphi}_j^{DR_\alpha}$ and $\hat{\varphi}_j^{EIR_\beta}$, by taking the average over all samples, M.

$$\hat{\varphi}_j^{DR_\alpha} = \frac{1}{M}\sum_m^M V_{m,j}^{DR_\alpha} \text{ and } \hat{\varphi}_j^{EIR_\beta} = \frac{1}{M}\sum_m^M V_{m,j}^{EIR_\beta}.$$

end
Notes: The integrals can be computed with quadrature methods. The algorithm for the balance of risks, BR, follows analogously.

4 Simulation Exercises

4.1 Model performance for different DGPs

This section describes the simulation exercise to evaluate the competing models' performance under different densities for $\{y_t\}_1^T$. Here, eight DGPs are considered:

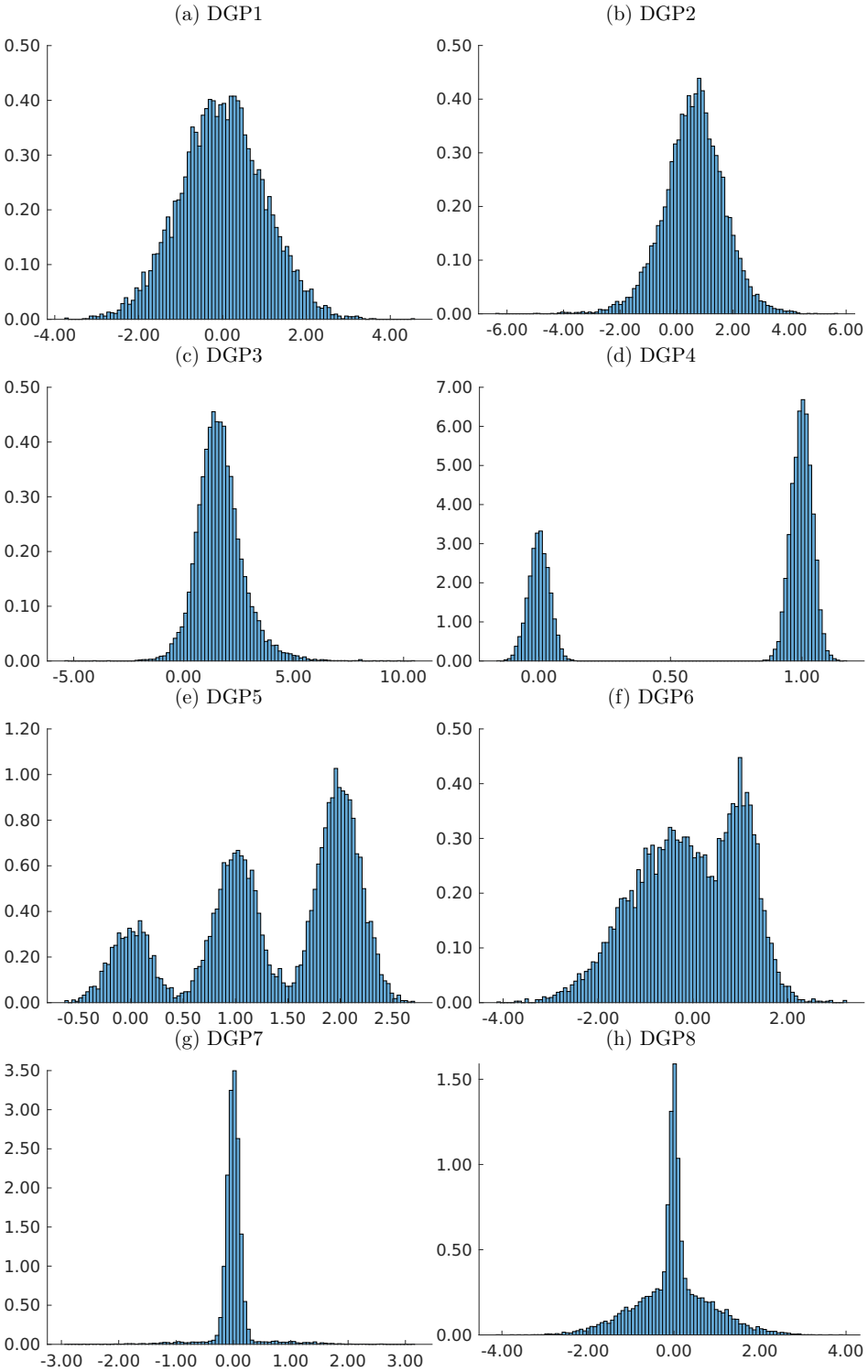
- Gaussian: $y \sim N(0, 1)$
- Negatively skewed: $y \sim ST(1, 1, -0.5, 10)$
- Skewed and Heavy-tailed: $y \sim ST(1, 1, 1, 5)$
- Bimodal, separate model: $y \sim \frac{1}{3}N(0, 0.04) + \frac{2}{3}N(1, 0.04)$
- Trimodal: $y \sim \frac{1}{6}N(0, 0.2) + \frac{2}{6}N(1, 0.2) + \frac{1}{2}N(2, 0.2)$

- Bimodal, skewed: $y \sim \frac{3}{4}N(\frac{-43}{100}, 1) + \frac{1}{4}N(\frac{107}{100}, \frac{1}{3})$
- Outlier: $y \sim \frac{1}{10}N(0, 1) + \frac{9}{10}N(1, 0.1)$
- Kurtotic: $y \sim \frac{2}{3}N(0, 1) + \frac{1}{3}N(1, 0.1)$,

where DGP1 to DGP5 follow Mitchell et al. (2024) and DGP6 to DGP8 are considered in several quantile regression papers (e.g. Korobilis and Schröder, 2025b; Lim et al., 2020). Histograms of the respective DGPs are given in Figure 1. The simulation exercise is repeated for $T = [50, 100, 200, 500]$. The performance is evaluated using the mean squared error for the first four moments as well as the Kullback-Leibler divergence, following Mitchell et al. (2024), over 500 Monte Carlo samples. In all cases, the density regression model allows for 5 mixture components.

The results are provided in Table 1. Overall, the models perform on par with the density regression model performing slightly better, particularly as the sample size increases. An exception is the outlier process in DGP7. While both models struggle to match the skewness and kurtosis of the underlying data, the errors are particularly large for the density regression model.

Figure 1: Histograms for simulated DGPs



Notes: The panels illustrate the histograms of the dependent variables under the different DGPs in the simulation exercise. All histograms are based on 10000 samples.

Table 1: Model performance for different DGPs

		Mean	Variance	Skewness	Kurtosis	KL	Mean	Variance	Skewness	Kurtosis	KL
T				DGP_1					GDP_2		
50	DR	0.00	0.00	0.02	0.85	0.03	0.00	0.01	0.10	6.44	0.04
	QR	0.00	0.02	0.05	0.57	0.04	0.00	0.03	0.12	1.67	0.05
100	DR	0.00	0.00	0.02	0.22	0.02	0.00	0.01	0.09	1.85	0.02
	QR	0.00	0.01	0.04	0.37	0.03	0.00	0.02	0.14	2.34	0.03
200	DR	0.00	0.00	0.01	0.07	0.01	0.00	0.00	0.06	0.64	0.01
	QR	0.00	0.00	0.03	0.18	0.02	0.00	0.01	0.08	1.09	0.02
500	DR	0.00	0.00	0.01	0.04	0.01	0.00	0.00	0.03	0.50	0.01
	QR	0.00	0.00	0.01	0.07	0.01	0.00	0.00	0.04	0.85	0.01
T				DGP_3					GDP_4		
50	DR	0.01	0.18	1.73	5087.19	0.06	0.00	0.00	0.01	0.00	0.87
	QR	0.00	0.13	0.60	12.97	0.06	0.00	0.00	0.01	0.06	2.25
100	DR	0.01	0.07	0.91	100.67	0.04	0.00	0.00	0.00	0.00	0.30
	QR	0.00	0.05	0.53	17.49	0.04	0.00	0.00	0.00	0.01	1.26
200	DR	0.00	0.03	0.70	22.29	0.03	0.00	0.00	0.00	0.00	0.08
	QR	0.00	0.04	0.54	27.83	0.03	0.00	0.00	0.00	0.01	0.92
500	DR	0.00	0.02	0.69	83.40	0.02	0.00	0.00	0.00	0.00	0.01
	QR	0.00	0.04	0.69	95.16	0.01	0.00	0.00	0.00	0.01	0.77
T				DGP_5					GDP_6		
50	DR	0.00	0.00	0.01	0.05	0.15	0.00	0.00	0.02	0.28	0.07
	QR	0.00	0.00	0.01	0.42	0.11	0.00	0.01	0.03	0.45	0.06
100	DR	0.00	0.00	0.00	0.00	0.16	0.00	0.00	0.01	0.08	0.05
	QR	0.00	0.00	0.01	0.22	0.12	0.00	0.01	0.02	0.19	0.05
200	DR	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.01	0.04	0.03
	QR	0.00	0.00	0.00	0.09	0.11	0.00	0.00	0.01	0.13	0.04
500	DR	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.02	0.02
	QR	0.00	0.00	0.00	0.02	0.08	0.00	0.00	0.01	0.05	0.02
T				DGP_7					GDP_8		
50	DR	0.00	0.00	15.80	128303.64	0.42	0.00	0.00	0.06	3.48	0.10
	QR	0.00	0.01	3.49	88.37	1.00	0.00	0.02	0.14	2.00	0.15
100	DR	0.00	0.00	6.91	937.14	0.27	0.00	0.00	0.05	0.74	0.07
	QR	0.00	0.00	5.13	219.85	0.45	0.00	0.01	0.13	1.22	0.11
200	DR	0.00	0.00	7.76	653.23	0.17	0.00	0.00	0.02	0.45	0.05
	QR	0.00	0.00	3.93	289.41	0.20	0.00	0.01	0.07	0.58	0.08
500	DR	0.00	0.01	10.58	665.76	0.09	0.00	0.00	0.01	0.36	0.04
	QR	0.00	0.01	2.14	375.35	0.11	0.00	0.00	0.04	0.30	0.05

The table contains the mean squared deviations of the estimated moments from the moments of the observed data as well as the Kullback-Leibler divergence. The individual panels provide the results for the different DGPs in Figure 1.

4.2 Mixture component selection with stick-breaking

This simulation builds on simulation exercise 4.1 and assess the accuracy of the stick breaking prior selecting the correct number of mixture components. To base the assessment on DGPs where the correct number of gaussian kernels is known, only DGP1 and DGP4 to DGP8 are considered. The maximum number of components is set to $\bar{C} = 5$.

Table 2: Number of Components selection for different DGPs

		DGP1	DGP4	DGP5	DGP6	DGP7	DGP8
50	Hitrate	0.19	1	0.81	0.73	0.67	0.24
	Median	2	2	3	2	2	3
100	Hitrate	0.32	1	0.99	0.82	0.51	0.21
	Median	2	2	3	2	2	3
200	Hitrate	0.55	1	0.99	0.89	0.36	0.15
	Median	1	2	3	2	3	3
500	Hitrate	0.77	1	0.94	0.95	0.29	0.1
	Median	1	2	3	2	3	3
	Simulated number	1	2	3	2	2	2

The table contains the hitrate and median estimated number of kernels. The hitrate measures the fraction of Monte Carlo iterations in which the correct number of Gaussian kernels is identified by the model.

As the Table 2 reveals, the stick breaking prior is particularly successful at identifying the correct number of mixture components for unimodal distributions as well as multimodal distributions with distinct modes. However, in case of the kurtotic and the outlier distribution, the number of components tends to be overestimated by one.

In applications, if one also wishes to find an optimal truncation point $\bar{C}$, the ELBO, i.e. the convergence criterion of the optimization problem can be used as it provides an approximation of the marginal likelihood. This gives rise to the following algorithm:

1. Set an upper bar on the number of mixture components $\bar{C}$.
2. For $C = 1, \dots, \bar{C}$, re-estimate the model with VB an store the ELBO.
3. Set
$$C^* = \arg\max_C ELBO_C.$$
4. Re-estimate/store the estimates for the model estimated with C^*.

Because the algorithm is fast, this optimization can even be repeated in real-time at every new data vintage, provided $\bar{C}$ is not too large. For macroeconomic data, $\bar{C} = 5$ or $\bar{C} = 10$ are useful starting points.

4.3 Performance metrics for decomposition algorithms

To evaluate the performance of the proposed decomposition algorithms this simulation exercise simulates a VAR with independent variables and a VAR with dependent variables, with three variables, a single lag, and no constants. The VAR with fully dependent endogenous variables is constructed with

$$\boldsymbol{Y}_t^D = \boldsymbol{B}^D \boldsymbol{Y}_{t-1}^D + \boldsymbol{u}_t^D, \quad \boldsymbol{u}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}^D),$$

where

$$\boldsymbol{B}^D = \begin{bmatrix} 0.85 & 0.10 & 0.40 \\ 0.10 & 0.80 & -0.20 \\ 0.30 & 0.50 & -0.60 \end{bmatrix} \text{ and } \boldsymbol{\Sigma}^D = \boldsymbol{C}^D \boldsymbol{C}^{D'}, \text{ with } \boldsymbol{C}^D = \begin{bmatrix} 0.20 & 0.30 & 0.10 \\ -0.40 & 0.50 & 0.20 \\ 0.30 & -0.20 & 0.60 \end{bmatrix}$$

and $\boldsymbol{Y}_t^D$ has dimensions $T \times N$. To ensure that the unconditional variances of the simulated dependent variables are the same in both cases, the VAR with independent variables is instead simulated according to

$$\boldsymbol{Y}_t^I = \boldsymbol{B}^I \boldsymbol{Y}_{t-1}^I + \boldsymbol{u}_t^I, \quad \boldsymbol{u}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}^I),$$

with $\boldsymbol{B}^I = diag(diag(\boldsymbol{B}^D))$ and $\boldsymbol{\Sigma}^I = \boldsymbol{C}^I \boldsymbol{C}^{I'}$, where $\boldsymbol{C}^I$ is derived according to the Lyapunov equation with $\boldsymbol{C}^I = diag(\Gamma \odot (\boldsymbol{1} - diag(\boldsymbol{B}^I)^2)$, and $\Gamma = diag(vec_{N\times N}^{-1}((\boldsymbol{I}_{N^2} - \boldsymbol{B}^D \bigotimes \boldsymbol{B}^D)^{-1} vec(\boldsymbol{\Sigma}^D)))$.

As alternatives to the sampling based decomposition algorithm proposed in the paper, the simulation proposes:

- The full Shapley value implementation which explores all coalition-time-combinations, denoted $SHAP_{Full}$, which are special cases of algorithm 2 and algorithm 3:

    1. For each sample of the background data, compute the marginal contribution of each variable in all possible coalitions of variables. For a given variable i and a specific coalition S (that does not contain variable i), the marginal contribution of i to coalition S is $M(i, S) = v(S \cup \{i\}) - v(S)$, with v denoting the objective function, i.e. either a probability bin of the forecast distribution or the risk measure.

2. Weight the marginal contributions and sum up these weighted marginal contributions for each variable.

- A kernel-based Shapley value variant, denoted $SHAP_{Kernel}$:

  1. For a given number of iterations, draw a random combination of background data instances and random permutation of variables to be replaced with background data samples.
  2. Assign a weight to each permutation depending on how many variables have been replaced.
  3. Estimate a weighted regression of the change in the output/prediction of the model onto indicators of whether a variable has not been replaced by background data.
  4. The coefficients then yield the approximate contributions.

- A conditional Shapley value variant based on kNN selected samples, denoted $SHAP_{KNN}$:

  1. Identify the local neighbourhood of the sample instance to be explained by means of kNN.
  2. Sample a random data point from this localized background data as well as a random permutation of the variables. Replace the value of the selected variables with the local background data sample.
  3. Compute the marginal contribution of each variable to each coalition and sum up the weighted marginal contributions to arrive at the final constribution estimate.

- Decompositions based on gradients, denoted $Gradients$:

  1. Compute the gradient of the model output/forecast at the data sample of interest, with respect to the intput data.
  2. Multiply the gradient with the difference between the current data point and the average across the background data to arrive at the final contribution estimate.

- Decompositions based on integrated gradients, denoted $I - Gradients$:

1. Define the number of steps (here 100). Then compute the corresponding stepsize for moves along the path from the current data sample to the average of the background data.
2. Compute and average the gradients along this path to arrive at the final estimate.

Each Monte Carlo iteration, the exercise cycles through the steps:

1. Simulate data according to the VAR dynamics.
2. Generate a density forecast/risk measure prediction for the first variable at the end of the sample. To abstract from estimation uncertainty, the forecast is generated based on the true parameters.
3. Decompose the forecast density/risk measure using the different decomposition algorithms.

Because no ground truth of the decompositions exists, performance is instead evaluated based on simple heuristics:

- **Error:**
  This metric evaluates the accuracy of the decomposition algorithm. It measures the difference between the sum of the variable's contributions and the change that is to be explained, i.e. the difference between the forecast and the average forecast constructed from the past data samples/the background data.

$$Error = \left\| \sum_{l=1}^{n} \varphi_l - [f_x(x)_{t+h|t} - E(f_x(x)_{t+h})] \right\|$$

  The larger Error, the less accurate the decomposition algorithm.

- **Infidelity Scores:**
  Measure the "unfaithfulness" of a model's reconstructed forecast, derived from decompositions, compared to its true forecast when using perturbed data. They are constructed as

$$Infidelity = \sqrt{\frac{1}{K} \sum_{k=1}^{K} \left[\Big((f_x(x)_{t+h|t}) - (f_{x_k^p}(x)_{t+h|t}) \Big) - \boldsymbol{\delta}_k \cdot \boldsymbol{\varphi} \right]^2},$$

where $\boldsymbol{x}_k^p = \boldsymbol{x}_t - \boldsymbol{\delta}_k$ with $\boldsymbol{\delta}_k \sim N(0,1)$ and dimensions $1 \times N$. $\boldsymbol{\varphi}$ stacks the variables' contributions. Thus, the further away the reconstructed forecast based on multiplying the size of the perturbation with the variables' attributions from the true forecast when using the perturbed data, the less "faithful" the decomposition.

- **Correlation:**
  This statistic is based on a simple heuristic. Given that the simulated data is Gaussian, a point forecast for $Y_{1,t+1}$ is given by $E(Y_{1,t+1}) = \sum_{i=1}^{N} B_{1,i}Y_{i,t}$. The variable for which $|B_{1,i}Y_{i,t}|$ is largest, hence has the largest impact on the location of $f_x(x)$. In the Gaussian context, a variable impacts the forecast distribution only through changes in the location or conditional mean. This then suggests that the largest aggregated changes in local probability mass should stem from the variable with the largest $B_{1,i}Y_{i,t}$, i.e. contribution to the mean. Across the support of the density, the absolute sum of the variables' local contributions to changes in probability mass as computed using any of the decomposition algorithms should then align with this ranking.

$$Correlation = Corr\left(|B_{1,i}Y_{i,t}|, \sum(|\boldsymbol{\varphi}|)\right),$$

  where $Corr$ denotes the Pearson correlation coefficient and $\sum(|\boldsymbol{\varphi}|)$ takes the sum across instances of the estimated contributions for each variable, e.g. across bins of the forecast distribution. In case of evaluating contributions to risk measures instead, the sum is replaced by the estimated variables' contributions.

- **Hitrate:**
  The hitrate follows the same logic as **Correlations**, however, instead it evaluates the fraction of times the correct variable is assigned the largest contribution as determined by $|B_{1,i}Y_{i,t}|$.

$$Hitrate = \frac{\sum_1^M \mathbf{1}[max_i(|\varphi_i|) = max_j(|B_{1,i}Y_{i,t}|_j)]}{M},$$

  where max_j denotes the largest element in each vector and M is the total number of Monte Carlo iterations.

The exercise is repeated for the density decompositions where a grid of 82 bins is assumed and for the risk measures with symmetric squared preferences. The results are

computed across 500 Monte Carlo iterations.

The results for the density decompositions are provided in Table 3. A few interesting patterns emerge. For the independent data, all competing decomposition algorithm identify the correct variable as the main contributor to changes in probability mass, already for very small sample sizes and have relatively similar infidelity scores. A key difference between the two groups of estimators, Shapely value-based and gradient-based, is that the former should aggregate to the difference between the forecast density and the forecast density constructed from the background data sample. As highlighted in the main body of the paper, this is an important feature, as it preserves integrability of the decomposed density and allows to compare contributions across vintages and forecast horizons. By construction, this property must hold for the exact method, however, the error remains negligible for the other variants. In contrast, both gradient-based methods produce decompositions that do not aggregate without error, which in case of IG can be mitigated by increasing the number of integration steps. The overall patterns are emphasized when moving to the dependent VAR. All metrics suggest that the Shapley value-based variants provide more accurate estimates, with smaller infidelity scores and noticeably larger hit rates and correlations. In addition, while the aggregation errors remain negligible for the Shapley value-based decompositions, they are inflated for the gradient-based methods.
Table 4 contains the corresponding results for the loss function decompositions. While infidelity scores and approximation errors (with the exception of the exact Shapley values) have now increased for all models, the overall patterns are preserved when data is independent. With dependent data, the aggregation errors of the Shapley value-based methods remain significantly smaller, while the other performance metrics are on par for both methods.

Table 3: Performance of different decomposition algorithms - Forecast density

	Error	Infidelity	Correlation	Hitrate	Error	Infidelity	Correlation	Hitrate
				Independent VAR				
		$T = 50$				$T = 100$		
SHAP_{Full}	0.00	0.03	1.00	1.00	0.00	0.03	1.00	1.00
$\text{SHAP}_{Sampling}$	0.00	0.03	1.00	1.00	0.00	0.03	1.00	1.00
SHAP_{Kernel}	0.01	0.03	1.00	1.00	0.01	0.03	1.00	1.00
SHAP_{KNN}	0.00	0.02	1.00	1.00	0.00	0.02	1.00	1.00
Gradients	0.34	0.04	1.00	1.00	0.36	0.04	1.00	1.00
I-Gradients	0.14	0.03	1.00	1.00	0.15	0.03	1.00	1.00
		$T = 200$				$T = 500$		
SHAP_{Full}	0.00	0.03	1.00	1.00	0.00	0.03	1.00	1.00
$\text{SHAP}_{Sampling}$	0.00	0.03	1.00	1.00	0.00	0.03	1.00	1.00
SHAP_{Kernel}	0.01	0.03	1.00	1.00	0.01	0.03	1.00	1.00
SHAP_{KNN}	0.00	0.02	1.00	1.00	0.00	0.02	1.00	1.00
Gradients	0.35	0.04	1.00	1.00	0.35	0.04	1.00	1.00
I-Gradients	0.15	0.03	1.00	1.00	0.15	0.03	1.00	1.00
				Dependent VAR				
		$T = 50$				$T = 100$		
SHAP_{Full}	0.00	0.05	0.83	0.86	0.00	0.05	0.84	0.87
$\text{SHAP}_{Sampling}$	0.00	0.05	0.83	0.86	0.00	0.05	0.84	0.87
SHAP_{Kernel}	0.06	0.05	0.82	0.86	0.07	0.05	0.84	0.87
SHAP_{KNN}	0.00	0.04	0.74	0.77	0.00	0.04	0.75	0.76
Gradients	0.73	0.09	0.74	0.80	0.89	0.10	0.75	0.78
I-Gradients	0.30	0.06	0.74	0.80	0.33	0.06	0.75	0.78
		$T = 200$				$T = 500$		
SHAP_{Full}	0.00	0.05	0.82	0.86	0.00	0.05	0.86	0.90
$\text{SHAP}_{Sampling}$	0.00	0.05	0.82	0.86	0.00	0.05	0.86	0.90
SHAP_{Kernel}	0.07	0.05	0.81	0.85	0.08	0.05	0.86	0.90
SHAP_{KNN}	0.00	0.04	0.77	0.79	0.01	0.04	0.82	0.83
Gradients	1.07	0.11	0.72	0.79	1.27	0.13	0.80	0.83
I-Gradients	0.37	0.07	0.72	0.79	0.39	0.07	0.80	0.83

The table contains different performance metrics for decomposing variables' contributions to probability bins of a forecast density. Error measures the accuracy of the attribution algorithm in terms of the contributions adding up to the difference between the benchmark and the forecast. Infidelity scores measure by how much a decomposition fails to explain the original forecast using only the estimated feature contributions. Correlation refers to the correlation between the sum of the estimated contributions across the densities support and the expected variable importance. The hitrate measures the number of times which the decomposition algorithm identifies the variable with the largest expected performance as the variable with the largest contribution.

Table 4: Performance of different decomposition algorithms - Risk measure

	Independent VAR							
	Error	Infidelity	Correlation	Hitrate	Error	Infidelity	Correlation	Hitrate
	$T = 50$				$T = 100$			
SHAP_{Full}	0.00	0.04	1.00	1.00	0.00	0.04	1.00	1.00
$\text{SHAP}_{Sampling}$	0.06	0.04	1.00	1.00	0.05	0.04	1.00	1.00
SHAP_{Kernel}	0.33	0.04	1.00	1.00	0.35	0.04	1.00	1.00
SHAP_{KNN}	0.05	0.02	1.00	1.00	0.06	0.02	1.00	1.00
Gradients	6.32	0.05	1.00	1.00	6.72	0.06	1.00	1.00
I-Gradients	2.68	0.03	1.00	1.00	2.86	0.03	1.00	1.00
	$T = 200$				$T = 500$			
SHAP_{Full}	0.00	0.04	1.00	1.00	0.00	0.04	1.00	1.00
$\text{SHAP}_{Sampling}$	0.03	0.04	1.00	1.00	0.02	0.04	1.00	1.00
SHAP_{Kernel}	0.39	0.04	1.00	1.00	0.40	0.04	1.00	1.00
SHAP_{KNN}	0.06	0.02	1.00	1.00	0.07	0.02	1.00	1.00
Gradients	6.94	0.06	1.00	1.00	7.75	0.06	1.00	1.00
I-Gradients	3.02	0.04	1.00	1.00	3.10	0.04	1.00	1.00
	Dependent VAR							
	$T = 50$				$T = 100$			
SHAP_{Full}	0.00	0.02	0.71	0.76	0.00	0.02	0.75	0.80
$\text{SHAP}_{Sampling}$	0.02	0.02	0.71	0.76	0.02	0.02	0.75	0.80
SHAP_{Kernel}	0.28	0.02	0.70	0.76	0.43	0.02	0.75	0.79
SHAP_{KNN}	0.02	0.02	0.65	0.74	0.03	0.02	0.64	0.75
Gradients	2.23	0.03	0.73	0.77	3.50	0.03	0.75	0.79
I-Gradients	0.84	0.02	0.73	0.77	1.26	0.02	0.75	0.79
	$T = 200$				$T = 500$			
SHAP_{Full}	0.00	0.03	0.76	0.80	0.00	0.03	0.78	0.82
$\text{SHAP}_{Sampling}$	0.02	0.03	0.76	0.80	0.02	0.03	0.78	0.82
SHAP_{Kernel}	0.50	0.03	0.76	0.80	0.65	0.03	0.78	0.82
SHAP_{KNN}	0.04	0.02	0.64	0.76	0.04	0.02	0.63	0.76
Gradients	4.59	0.04	0.77	0.80	5.66	0.04	0.79	0.82
I-Gradients	1.85	0.03	0.77	0.80	2.39	0.02	0.79	0.82

The table contains different performance metrics for decomposing variables' contributions to the risk measure with squared symmetric preferences. Error measures the accuracy of the attribution algorithm in terms of the contributions adding up to the difference between the benchmark and the forecast. Infidelity scores measure by how much a decomposition fails to explain the original forecast using only the estimated feature contributions. Correlation refers to the correlation between the sum of the estimated contributions across the densities support and the expected variable importance. The hitrate measures the number of times which the decomposition algorithm identifies the variable with the largest expected performance as the variable with the largest contribution.

4.4 Decomposition performance for different models

This forecasting exercise simulates a simple DGP to demonstrate that the decomposition algorithm for the forecast density is accurate irrespective of using the density regression or quantile regression model. y_t is constructed as

$$y_t = 0.5 + 0.8 \cdot y_{t-1} + \boldsymbol{\beta} \cdot \boldsymbol{x}_t + u_t, \text{ with } \quad u_t \sim N(0, \sigma), \text{ and } \quad \boldsymbol{x}_t \sim N(\mathbf{0}, \mathbf{1}),$$

where $\boldsymbol{\beta} = [0.6, 0.5, -0.2]$ and $\sigma = 2$. For simplicity, in this exercise the density is discretized over a grid of six bins and decompositions are computed using the sampling-based approach. The performance metrics are computed over 500 Monte Carlo iterations.

Table 5: Decomposition performance for different models - Densities

	Error	Infidelity	Correlation	Hitrate	Error	Infidelity	Correlation	Hitrate
		$T = 50$				$T = 100$		
DR	0.01	0.03	0.76	1.00	0.01	0.03	0.76	1.00
QR	0.00	0.02	0.69	1.00	0.01	0.02	0.71	1.00
		$T = 200$				$T = 500$		
DR	0.01	0.02	0.77	1.00	0.01	0.02	0.77	1.00
QR	0.01	0.02	0.75	1.00	0.01	0.02	0.77	1.00

The table contains different performance metrics for decomposing variables' contributions to probability bins of a forecast density for the density regression and the quantile regression model. Error measures the accuracy of the attribution algorithm in terms of the contributions adding up to the difference between the benchmark and the forecast. Infidelity scores measure by how much a decomposition fails to explain the original forecast using only the estimated feature contributions. Correlation refers to the correlation between the sum of the estimated contributions across the densities support and the expected variable importance. The hitrate measures the number of times which the decomposition algorithm identifies the variable with the largest expected performance as the variable with the largest contribution.

Table 6: Decomposition performance for different models -Risk measures

	Error	Infidelity	Correlation	Hitrate	Error	Infidelity	Correlation	Hitrate
		$T = 50$				$T = 100$		
DR	0.61	9.61	0.70	1.00	0.64	9.28	0.73	1.00
QR	0.43	6.18	0.68	1.00	0.55	7.21	0.71	1.00
		$T = 200$				$T = 500$		
DR	0.64	9.28	0.73	1.00	0.69	8.40	0.75	1.00
QR	0.57	8.08	0.72	1.00	0.59	7.55	0.76	1.00

The table contains different performance metrics for decomposing variables' contributions to probability bins of a forecast density for the density regression and the quantile regression model. Error measures the accuracy of the attribution algorithm in terms of the contributions adding up to the difference between the benchmark and the forecast. Infidelity scores measure by how much a decomposition fails to explain the original forecast using only the estimated feature contributions. Correlation refers to the correlation between the sum of the estimated contributions across the densities support and the expected variable importance. The hitrate measures the number of times which the decomposition algorithm identifies the variable with the largest expected performance as the variable with the largest contribution.

The results, given in Table 5 and Table 6, suggest that for small sample sizes the density regression model produces slightly less accurate decompositions in terms of error and infidelity scores, but more accurate decompositions in terms of the correlation metric. As the sample size increases, the accuracy converges. In addition, in all cases the models identify the largest contributor correctly. The inflated metrics in case of the risk measure application are largely due to the risk measure's scale. The error can generally be decreased by either relying on the exact implementation or by increasing the number of samples in the sampling-based procedure.

5 Additional Results

This section presents additional result tables. Table 7, Table 9, and Table 8 in subsection 5.1 contain density calibration tests, quantile scores and CRPS for headline PCE, headline CPI, and core CPI inflation, respectively, corresponding to the exercise in section 4.2 of the main paper.

Table 10, Table 12, Table 11, Table 13 provide the test results on statistically significant differences in quantile scores and CRPS of the density regression and quantile regression modes vis-a-vis the UC-SV, following Manzan (2015).

Table 14 in subsection 5.2 provides point-forecast performance statistics and Diebold-Mariano tests for core and headline PCE and CPI inflation.

5.1 Density Forecast Performance

Table 7: Density specification tests - PCE inflation

	Density specification tests															
	full		center		left tail		right tail		full		center		left tail		right tail	
	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM
	$H=1$								$H=2$							
DR	**1.54**	**0.78**	**1.54**	**1.35**	0.92	0.24	0.79	0.18	**1.45**	**0.65**	1.45	**1.02**	**0.92**	0.17	**1.35**	**0.40**
QR	1.05	0.27	1.05	0.29	0.93	0.16	0.9	**0.35**	**1.79**	**0.86**	1.26	0.55	**0.75**	0.14	**1.79**	**2.20**
UC-SV	**1.76**	**0.95**	**1.76**	**1.16**	**1.40**	**0.73**	**1.60**	**0.76**	**1.58**	**0.66**	**1.58**	**0.90**	**1.33**	**0.60**	0.75	0.27
	$H=3$								$H=4$							
DR	0.89	0.27	0.89	0.30	0.63	0.12	0.81	**0.35**	1.42	0.60	1.42	0.61	0.76	0.16	**1.33**	**1.01**
QR	**2.34**	**1.41**	**1.84**	**0.97**	**0.78**	0.14	**2.34**	**3.56**	**2.58**	**1.98**	**2.29**	**1.25**	**0.93**	0.22	**2.58**	**5.19**
UC-SV	**1.66**	**0.73**	**1.66**	**1.01**	**1.43**	**0.62**	0.84	0.27	**1.72**	**0.68**	**1.72**	**1.07**	**1.22**	**0.53**	0.61	0.07
	$H=8$								$H=12$							
DR	1.12	0.17	0.44	0.04	0.35	0.03	1.12	0.55	1.71	**0.96**	**1.71**	**1.18**	0.31	0.02	**1.55**	**1.46**
QR	1.40	0.34	0.71	0.09	**0.86**	0.17	**1.40**	**1.02**	0.60	0.05	0.40	0.02	0.58	0.08	0.60	0.09
UC-SV	1.30	0.30	1.30	0.44	0.76	**0.26**	0.56	0.08	1.36	0.47	1.30	0.60	**1.36**	**0.61**	0.54	0.07

	Quantile scores							CRPS	*Quantile scores*							CRPS
	0.05	0.1	0.2	0.5	0.8	0.9	0.95		0.5	0.1	0.2	0.5	0.8	0.9	0.95	
	$H=1$								$H=2$							
DR	-0.21	-0.30	-0.42	-0.53	-0.35	-0.22	-0.14	0.77	-0.31	-0.42	-0.57	-0.72	-0.51	-0.35	-0.21	1.04
QR	-0.27	-0.35	-0.37	-0.59	-0.46	-0.30	-0.18	0.86	-0.28	-0.36	-0.50	-0.69	-0.54	-0.42	-0.28	1.03
UC-SV	-0.24	-0.37	-0.51	-0.56	-0.46	-0.33	-0.20	0.90	-0.26	-0.40	-0.58	-0.63	-0.46	-0.31	-0.19	0.98
	$H=3$								$H=4$							
DR	-0.30	-0.39	-0.53	-0.71	-0.56	-0.40	-0.28	1.04	-0.32	-0.43	-0.60	-0.81	-0.67	-0.51	-0.38	1.19
QR	-0.28	-0.38	-0.50	-0.75	-0.63	-0.48	-0.31	1.12	-0.29	-0.38	-0.52	-0.80	-0.63	-0.45	-0.33	1.14
UC-SV	-0.26	-0.41	-0.58	-0.65	-0.48	-0.34	-0.21	1.01	-0.26	-0.41	-0.59	-0.68	-0.55	-0.38	-0.24	1.08
	$H=8$								$H=12$							
DR	-0.29	-0.40	-0.54	-0.78	-0.73	-0.62	-0.53	1.21	-0.31	-0.44	-0.67	-0.87	-0.69	-0.52	-0.39	1.29
QR	-0.28	-0.40	-0.51	-0.79	-0.70	-0.45	-0.33	1.15	-0.29	-0.39	-0.49	-0.78	-0.65	-0.37	-0.31	1.11
UC-SV	-0.25	-0.40	-0.59	-0.82	-0.69	-0.50	-0.34	1.26	-0.26	-0.41	-0.59	-0.77	-0.66	-0.46	-0.30	1.20

The top half of the table contains the test statistics for the density specification tests in Rossi and Sekhposyan (2019). KS denotes the Kolmogorov-Smirnov and CVM the Kramér-von Mises based test statistic. Bold values indicate that the null hypothesis of correct specification can be rejected at 5% significance level. The bottom part of the table contains quantile scores following Manzan (2015), where smaller values indicate better performance.

Table 8: Density specification tests - CPI excluding food and energy inflation

	Density specification tests															
	full		center		left tail		right tail		full		center		left tail		right tail	
	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM
	$H=1$								$H=2$							
DR	0.85	0.11	0.85	0.15	0.70	0.11	0.45	0.03	0.87	0.21	0.82	0.21	**0.72**	**0.21**	**0.87**	**0.20**
QR	**1.89**	**1.27**	**1.74**	**1.37**	0.75	0.13	**1.89**	**2.21**	**2.56**	**2.15**	**2.31**	**1.97**	**1.07**	**0.35**	**2.56**	**4.32**
UC-SV	**2.83**	**3.05**	**2.83**	**4.51**	**2.25**	**1.66**	**2.25**	**1.55**	**2.57**	**2.53**	**2.57**	**3.55**	**2.25**	**1.71**	**1.84**	**1.33**
	$H=3$								$H=4$							
DR	0.55	0.03	0.55	0.03	**0.42**	**0.05**	0.48	0.04	0.90	0.23	0.79	0.19	**0.90**	**0.35**	0.85	0.21
QR	**2.54**	**1.80**	**2.23**	**1.66**	**1.23**	**0.58**	**2.54**	**3.31**	**2.57**	**2.36**	**2.52**	**2.40**	**1.16**	**0.47**	**2.57**	**4.16**
UC-SV	**2.62**	**2.38**	**2.62**	**3.37**	**2.25**	**1.67**	**1.76**	**1.12**	**2.38**	**2.00**	**2.38**	**2.70**	**2.25**	**1.70**	**1.53**	**0.91**
	$H=8$								$H=12$							
DR	1.17	0.37	1.17	0.54	**0.84**	**0.21**	0.81	0.18	1.39	0.34	0.89	0.18	0.33	0.03	**1.39**	**0.98**
QR	1.32	0.52	1.31	0.38	**1.21**	**0.65**	**1.32**	**0.65**	1.32	0.43	1.32	0.53	**1.02**	**0.33**	1.07	0.33
UC-SV	**2.19**	**1.43**	**2.19**	**1.85**	**2.04**	**1.55**	**1.26**	0.46	**2.11**	**1.28**	**2.11**	**1.68**	**2.04**	**1.56**	0.90	0.21

	Quantile scores							CRPS	*Quantile scores*							CRPS
	0.05	0.1	0.2	0.5	0.8	0.9	0.95		0.5	0.1	0.2	0.5	0.8	0.9	0.95	
	$H=1$								$H=2$							
DR	-0.12	-0.17	-0.25	-0.35	-0.28	-0.22	-0.18	0.53	-0.12	-0.18	-0.27	-0.38	-0.32	-0.27	-0.22	0.59
QR	-0.16	-0.17	-0.21	-0.29	-0.27	-0.21	-0.19	0.46	-0.16	-0.18	-0.24	-0.35	-0.33	-0.28	-0.23	0.55
UC-SV	-0.19	-0.30	-0.40	-0.28	-0.40	-0.30	-0.19	0.64	-0.19	-0.30	-0.40	-0.31	-0.39	-0.28	-0.18	0.65
	$H=3$								$H=4$							
DR	-0.19	-0.27	-0.37	-0.53	-0.50	-0.44	-0.32	0.78	-0.20	-0.28	-0.37	-0.48	-0.38	-0.30	-0.25	0.73
QR	-0.17	-0.20	-0.25	-0.36	-0.38	-0.33	-0.28	0.60	-0.18	-0.21	-0.26	-0.39	-0.44	-0.38	-0.31	0.65
UC-SV	-0.19	-0.30	-0.41	-0.34	-0.40	-0.29	-0.17	0.67	-0.19	-0.31	-0.42	-0.39	-0.44	-0.32	-0.21	0.73
	$H=8$								$H=12$							
DR	-0.26	-0.33	-0.43	-0.53	-0.46	-0.37	-0.31	0.86	-0.20	-0.28	-0.41	-0.62	-0.59	-0.49	-0.40	0.93
QR	-0.18	-0.25	-0.29	-0.44	-0.49	-0.43	-0.38	0.73	-0.18	-0.25	-0.29	-0.50	-0.49	-0.40	-0.41	0.74
UC-SV	-0.20	-0.31	-0.44	-0.48	-0.47	-0.33	-0.20	0.82	-0.20	-0.31	-0.44	-0.51	-0.50	-0.34	-0.20	0.85

The top half of the table contains the test statistics for the density specification tests in Rossi and Sekhposyan (2019). KS denotes the Kolmogorov-Smirnov and CVM the Kramér-von Mises based test statistic. Bold values indicate that the null hypothesis of correct specification can be rejected at 5% significance level. The bottom part of the table contains quantile scores following Manzan (2015), where smaller values indicate better performance.

Table 9: Density specification tests - CPI inflation

	Density specification tests															
	full		center		left tail		right tail		full		center		left tail		right tail	
	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM	KS	CVM
	$H=1$								$H=2$							
DR	1.16	0.40	1.12	0.52	0.59	0.10	**1.16**	**0.45**	1.23	0.34	1.12	0.45	0.38	0.04	**1.23**	**0.44**
QR	1.14	0.29	1.11	0.34	0.43	0.05	**1.14**	**0.43**	**1.69**	**1.03**	**1.67**	**0.93**	0.63	0.13	**1.69**	**2.14**
UC-SV	**1.64**	**0.82**	**1.64**	**0.91**	**1.43**	**0.58**	**1.59**	**0.90**	**1.55**	**0.63**	**1.55**	**0.84**	**1.25**	**0.44**	0.96	**0.38**
	$H=3$								$H=4$							
DR	1.28	**0.59**	1.28	0.74	0.61	0.08	**1.26**	**0.80**	**2.21**	**1.65**	**2.21**	**2.10**	0.66	0.10	**2.00**	**2.31**
QR	**2.31**	**1.52**	**2.00**	**1.23**	**0.79**	0.15	**2.31**	**3.48**	**2.43**	**1.74**	**1.98**	**1.33**	**0.82**	0.23	**2.43**	**4.08**
UC-SV	**1.47**	**0.65**	1.47	0.80	**1.42**	**0.54**	**1.06**	**0.47**	**1.69**	**0.60**	**1.69**	0.91	**1.13**	**0.46**	0.61	0.11
	$H=8$								$H=12$							
DR	1.06	0.43	1.06	0.58	**0.78**	0.17	0.86	0.40	1.36	0.44	0.96	0.45	0.47	0.04	**1.36**	**0.82**
QR	1.33	0.44	1.01	0.19	**0.95**	**0.42**	**1.33**	**0.94**	0.99	0.19	0.99	0.20	**0.85**	**0.26**	0.60	0.10
UC-SV	1.00	0.26	1.00	0.38	**0.92**	**0.24**	0.47	0.03	1.33	0.41	1.33	0.60	**1.12**	**0.37**	0.61	0.09

	Quantile scores							CRPS	*Quantile scores*							CRPS
	0.05	0.1	0.2	0.5	0.8	0.9	0.95		0.5	0.1	0.2	0.5	0.8	0.9	0.95	
	$H=1$								$H=2$							
DR	-0.32	-0.45	-0.61	-0.73	-0.58	-0.41	-0.28	1.10	-0.37	-0.49	-0.64	-0.79	-0.61	-0.42	-0.28	1.17
QR	-0.36	-0.45	-0.62	-0.81	-0.56	-0.39	-0.26	1.19	-0.39	-0.55	-0.68	-0.92	-0.78	-0.56	-0.39	1.38
UC-SV	-0.33	-0.50	-0.70	-0.76	-0.59	-0.42	-0.26	1.21	-0.35	-0.55	-0.77	-0.85	-0.58	-0.38	-0.24	1.30
	$H=3$								$H=4$							
DR	-0.43	-0.56	-0.74	-0.89	-0.67	-0.45	-0.29	1.32	-0.47	-0.61	-0.80	-1.03	-0.83	-0.62	-0.43	1.54
QR	-0.40	-0.55	-0.68	-0.94	-0.82	-0.62	-0.40	1.42	-0.40	-0.55	-0.67	-0.93	-0.77	-0.62	-0.49	1.43
UC-SV	-0.35	-0.55	-0.77	-0.88	-0.59	-0.40	-0.24	1.32	-0.36	-0.55	-0.78	-0.88	-0.68	-0.45	-0.28	1.39
	$H=8$								$H=12$							
DR	-0.44	-0.59	-0.75	-0.96	-0.79	-0.61	-0.46	1.49	-0.42	-0.62	-0.83	-1.02	-0.82	-0.60	-0.40	1.59
QR	-0.41	-0.58	-0.69	-0.93	-0.90	-0.59	-0.54	1.48	-0.41	-0.58	-0.69	-0.92	-0.78	-0.51	-0.40	1.40
UC-SV	-0.34	-0.53	-0.79	-1.04	-0.81	-0.58	-0.39	1.57	-0.35	-0.54	-0.77	-0.96	-0.81	-0.57	-0.37	1.51

The top half of the table contains the test statistics for the density specification tests in Rossi and Sekhposyan (2019). KS denotes the Kolmogorov-Smirnov and CVM the Kramér-von Mises based test statistic. Bold values indicate that the null hypothesis of correct specification can be rejected at 5% significance level. The bottom part of the table contains quantile scores following Manzan (2015), where smaller values indicate better performance.

Table 10: Significance of differences in quantile scores, PCE excluding energy and food

	Quantile scores							CRPS	*Quantile scores*							CRPS
	0.05	0.1	0.2	0.5	0.8	0.9	0.95		0.5	0.1	0.2	0.5	0.8	0.9	0.95	
	$H = 1$								$H = 2$							
DR	**2.63**	**3.75**	**3.43**	-0.69	1.48	0.91	-0.03	-1.34	-0.01	1.22	1.90	-1.28	0.56	0.49	-0.24	-0.68
QR	1.63	**4.50**	**4.55**	-0.40	**2.39**	**2.23**	1.38	**-2.23**	1.70	**3.04**	**3.41**	-0.73	0.70	1.64	0.57	-1.08
	$H = 3$								$H = 4$							
DR	-0.21	0.71	1.23	-1.50	0.59	0.13	-0.47	-0.19	-1.17	-0.27	0.56	-1.06	0.82	0.60	0.02	-0.33
QR	1.68	**3.69**	**3.35**	-0.99	-0.08	-0.75	-1.27	-0.31	1.73	**2.98**	**3.66**	-0.91	-0.28	-0.63	-1.07	-0.28
	$H = 8$								$H = 12$							
DR	-1.86	-1.38	-1.02	-1.43	-0.53	-0.78	-0.99	0.94	-0.11	0.36	-0.15	-1.09	-0.57	-1.22	**-2.01**	0.58
QR	1.95	**2.91**	**3.31**	0.22	-1.18	-0.41	-0.88	-0.65	**1.97**	**2.93**	**3.52**	-0.31	-1.10	-1.49	-1.84	-0.36

Entries in this table are t-statistic values for the null hypothesis of equal accuracy of the quantile forecasts from the density regression model and the quantile regression model relative to the UC-SV benchmark. Negative values indicate better performance of the former and positive values better performance of the latter. Bold values indicate statistically significant performance differences at the 5% level.

Table 11: Significance of differences in quantile scores, PCE

	Quantile scores							CRPS	*Quantile scores*							CRPS
	0.05	0.1	0.2	0.5	0.8	0.9	0.95		0.5	0.1	0.2	0.5	0.8	0.9	0.95	
	$H = 1$								$H = 2$							
DR	0.35	0.77	1.04	0.34	**2.04**	**2.70**	**2.41**	-1.08	-0.58	-0.16	0.11	-0.82	-0.75	-0.69	-0.46	0.35
QR	-0.33	0.14	1.36	-0.32	0.04	0.60	0.76	-0.31	-0.27	0.45	0.77	-0.53	-1.12	-1.59	-1.57	0.29
	$H = 3$								$H = 4$							
DR	-0.43	0.13	0.54	-0.53	-1.07	-1.00	-1.06	0.23	-0.61	-0.18	-0.06	-1.20	-1.27	-1.44	-1.72	0.70
QR	-0.27	0.30	0.87	-0.91	-1.61	-1.75	-1.47	0.63	-0.28	0.33	0.73	-1.13	-0.86	-0.79	-1.10	0.35
	$H = 8$								$H = 12$							
DR	-0.46	0.02	0.57	0.38	-0.29	-0.88	-1.42	-0.24	-0.53	-0.26	-0.78	-0.91	-0.28	-0.62	-1.11	0.55
QR	-0.43	-0.03	0.84	0.33	-0.06	0.53	0.07	-0.59	-0.30	0.19	0.94	-0.13	0.16	1.14	-0.13	-0.52

Entries in this table are t-statistic values for the null hypothesis of equal accuracy of the quantile forecasts from the density regression model and the quantile regression model relative to the UC-SV benchmark. Negative values indicate better performance of the former and positive values better performance of the latter. Bold values indicate statistically significant performance differences at the 5% level.

Table 12: Significance of differences in quantile scores, CPI excluding energy and food

	Quantile scores							CRPS	*Quantile scores*							CRPS
	0.05	0.1	0.2	0.5	0.8	0.9	0.95		0.5	0.1	0.2	0.5	0.8	0.9	0.95	
				$H = 1$								$H = 2$				
DR	**2.34**	**3.41**	**3.09**	-1.20	1.82	1.20	0.12	-1.14	**2.04**	**3.17**	**2.95**	-1.02	0.87	0.24	-0.60	-0.69
QR	0.78	**3.25**	**4.15**	-0.32	**2.18**	1.42	-0.03	**-1.97**	0.71	**2.99**	**3.61**	-0.56	0.86	0.02	-0.83	-1.15
				$H = 3$								$H = 4$				
DR	-0.02	0.71	0.67	**-2.03**	-0.86	-1.20	-1.47	0.82	-0.16	0.65	0.76	-1.24	0.62	0.17	-0.54	0.02
QR	0.49	**2.56**	**3.66**	-0.32	0.17	-0.60	-1.38	-0.72	0.37	**2.30**	**3.35**	0.01	-0.06	-0.69	-1.30	-0.68
				$H = 8$								$H = 12$				
DR	-1.22	-0.32	0.12	-0.52	0.09	-0.51	-1.29	0.33	-0.01	0.73	0.44	-1.27	-0.95	-1.60	**-2.25**	0.61
QR	0.46	1.51	**2.83**	0.60	-0.24	-1.24	**-2.12**	-0.74	0.34	1.40	**2.92**	0.15	0.14	-0.82	**-2.79**	-0.92

Entries in this table are t-statistic values for the null hypothesis of equal accuracy of the quantile forecasts from the density regression model and the quantile regression model relative to the UC-SV benchmark. Negative values indicate better performance of the former and positive values better performance of the latter. Bold values indicate statistically significant performance differences at the 5% level.

Table 13: Significance of differences in quantile scores, CPI

	Quantile scores							CRPS	*Quantile scores*							CRPS
	0.05	0.1	0.2	0.5	0.8	0.9	0.95		0.5	0.1	0.2	0.5	0.8	0.9	0.95	
				$H = 1$								$H = 2$				
DR	0.13	0.40	0.72	0.22	0.14	0.16	-0.35	-0.58	-0.17	0.41	0.93	0.42	-0.35	-0.50	-0.87	-0.64
QR	-0.26	0.38	0.62	-0.39	0.33	0.44	-0.05	-0.10	-0.30	-0.03	0.65	-0.52	-1.67	-1.74	**-2.01**	0.38
				$H = 3$								$H = 4$				
DR	-0.62	-0.09	0.25	-0.10	-0.92	-0.75	-1.00	0.02	-0.80	-0.38	-0.12	-1.00	-1.23	-1.52	-1.53	0.67
QR	-0.34	-0.05	0.64	-0.39	**-2.00**	**-2.10**	**-2.08**	0.46	-0.28	-0.01	0.76	-0.32	-0.73	-1.53	**-2.14**	0.18
				$H = 8$								$H = 12$				
DR	-0.90	-0.45	0.28	0.57	0.12	-0.24	-0.52	-0.37	-0.61	-0.57	-0.40	-0.45	-0.14	-0.30	-0.34	0.37
QR	-0.58	-0.35	0.66	0.80	-0.61	-0.06	-1.23	-0.37	-0.44	-0.32	0.56	0.25	0.26	0.59	-0.34	-0.48

Entries in this table are t-statistic values for the null hypothesis of equal accuracy of the quantile forecasts from the density regression model and the quantile regression model relative to the UC-SV benchmark. Negative values indicate better performance of the former and positive values better performance of the latter. Bold values indicate statistically significant performance differences at the 5% level.

5.2 Point Forecast Performance

Table 14: Relative RMSE vis-à-vis UC-SV

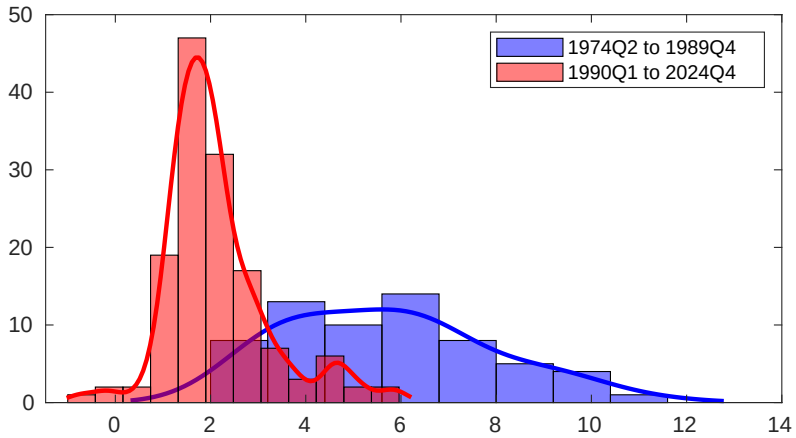
	PCE excluding energy and food											
	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
DR	1.27	1.16***	1.23**	1.06	1.11	1.09	1.03	1.19***	1.04	0.99	1.08	1.21
QR	1.12*	1.11**	1.16**	1.16*	1.13*	1.04*	1.01	0.98	0.98	1.00	0.97	1.00
	PCE											
	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
DR	0.87	1.09	1.04	1.08	1.06	0.98	0.94	0.99	0.99	0.99	1.06	1.03
QR	1.01	1.11**	1.17**	1.09	1.09	0.97	0.92	0.94	0.94	0.99	1.01	1.00
	CPI excluding energy and food											
	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
DR	1.38	1.31*	1.73**	1.20**	1.26**	1.20***	1.24***	1.20***	1.11**	1.18	1.09	1.13**
QR	1.07	1.18	1.18	1.04	1.05	1.02	0.99	0.96	0.97	0.97	1.00	0.97
	CPI											
	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
DR	0.96	0.96	1.00	1.08*	0.97	0.92	0.93*	0.94	0.92**	1.01	1.07	1.02
QR	1.04	1.08	1.10	0.99	0.95	0.93*	0.92**	0.92*	0.93*	0.94	0.95	0.94

The table contains the RMSE relative to the UC-SV model. Statistical significance of the Diebold-Mariano test is indicated by the $*$, where $\{***\} = 1\%$, $\{**\} = 5\%$, $\{*\} = 10\%$. For the QR model, RMSE are computed based on forecasts for the median.

5.3 Additional Figures

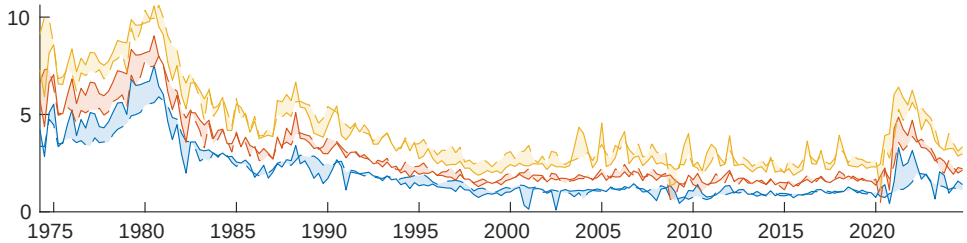
This section provides additional figures, referenced in the main body of the paper. Figure 2 provides illustrative evidence on the non-standard unconditional distribution of PCE core inflation, Figure 3 shows a time series for the entire sample of estimated quantile levels for the density regression and quantile regression model. Figure 4 provides the balance of risk for PCE core inflation for $h = 1$ and Figure 5, Figure 7, and Figure 6 provide the balance of risks for headline PCE and CPI, as well as core CPI inflation for $h = 4$. In turn, Figure 8, Figure 10, and Figure 9 show the decompositions of the balance of risk for headline PCE and CPI, as well as core CPI inflation.

Figure 2: Histograms of Annualized Quarterly Inflation



Notes: The histograms display the observed values for annualized quarterly PCE excluding energy and food inflation from 1974Q1 to 1889Q4 (blue) and from 1990Q1 to 2024Q4 (red). The corresponding lines show kernel density estimates over these sub-samples.

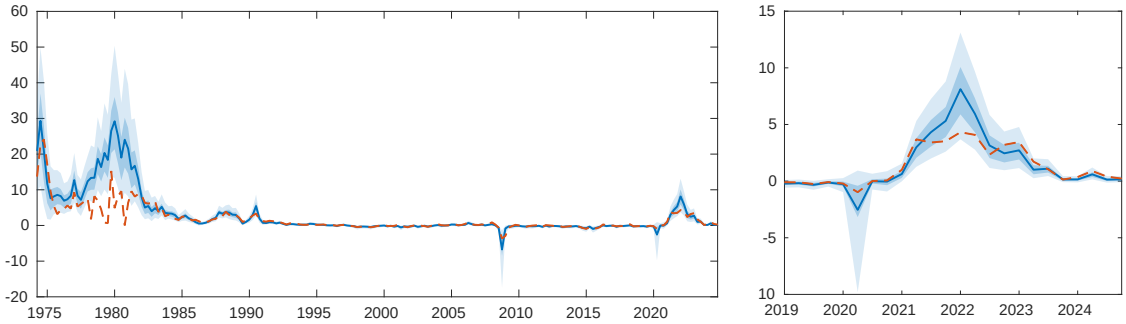
Figure 3: Estimated conditional quantiles



Notes: The chart shows the conditional quantiles -10th (blue), 50th (red), and 90th (yellow) - for the density regression model (solid lines) and the quantile regression model (dashed lines).

Figure 4: The Balance of Risk - PCE excluding energy and food

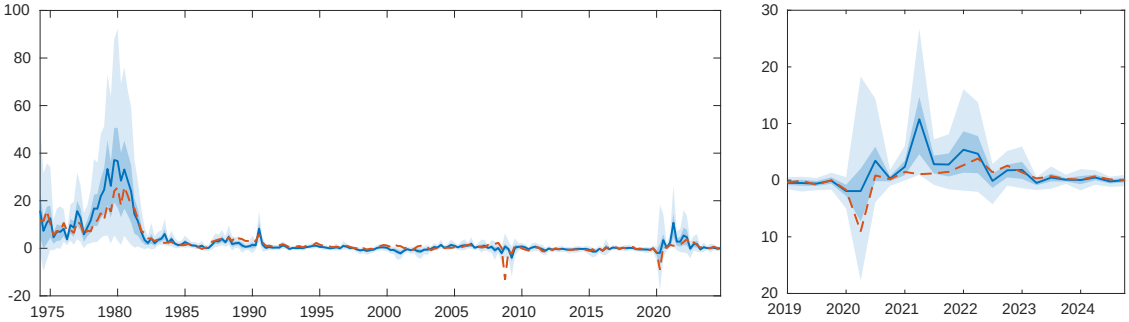
(a) BR for $\alpha = \beta = 2$ and $h = 1$ (b) BR 2019Q1 to 2024Q4



Notes: The solid blue line indicates the estimates for the density regression model over the full sample obtained. The dark shaded area corresponds to the 50% and the light shaded area to the 90% credible interval. The dashed red line shows the corresponding estimates computed with the quantile regression model.

Figure 5: The Balance of Risk - PCE

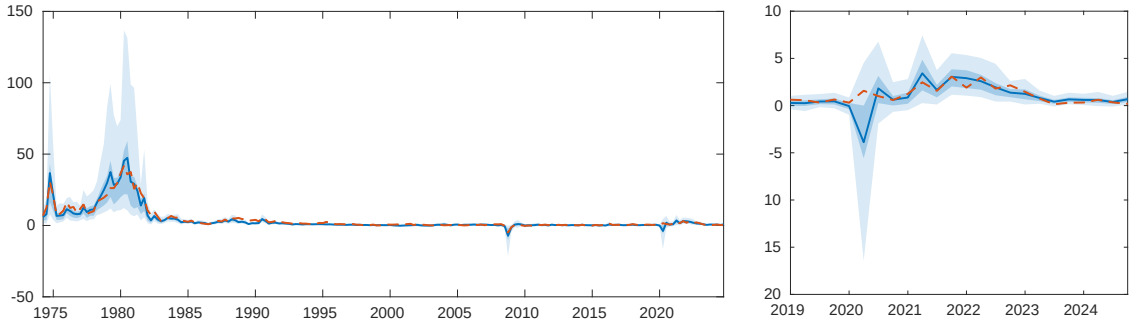
(a) BR for $\alpha = \beta = 2$ and $h = 4$ (b) BR 2019Q1 to 2024Q4



Notes: The solid blue line indicates the estimates for the density regression model over the full sample obtained. The dark shaded area corresponds to the 50% and the light shaded area to the 90% credible interval. The dashed red line shows the corresponding estimates computed with the quantile regression model.

Figure 6: The Balance of Risk - CPI excluding energy and food

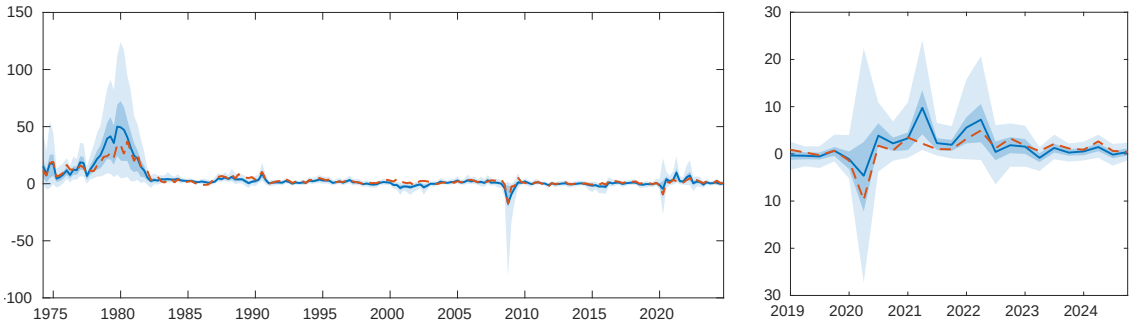
(a) BR for $\alpha = \beta = 2$ and $h = 4$ (b) BR 2019Q1 to 2024Q4



Notes: The solid blue line indicates the estimates for the density regression model over the full sample obtained. The dark shaded area corresponds to the 50% and the light shaded area to the 90% credible interval. The dashed red line shows the corresponding estimates computed with the quantile regression model.

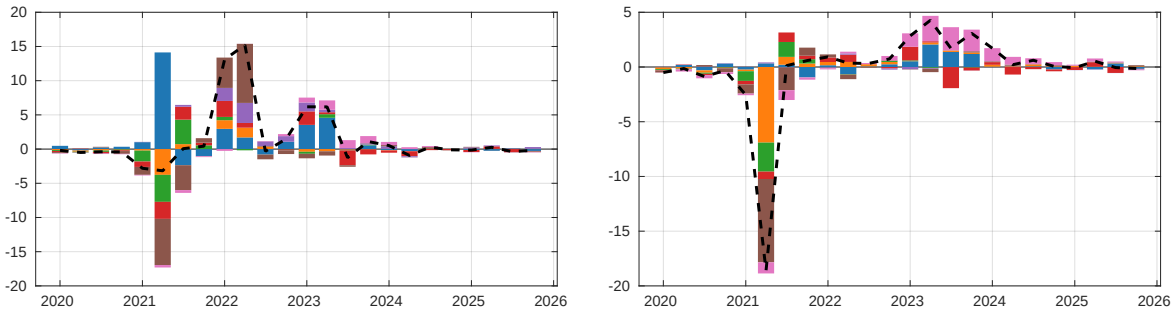
Figure 7: The Balance of Risk - CPI

(a) BR for $\alpha = \beta = 2$ and $h = 4$ (b) BR 2019Q1 to 2024Q4



Notes: The solid blue line indicates the estimates for the density regression model over the full sample obtained. The dark shaded area corresponds to the 50% and the light shaded area to the 90% credible interval. The dashed red line shows the corresponding estimates computed with the quantile regression model.

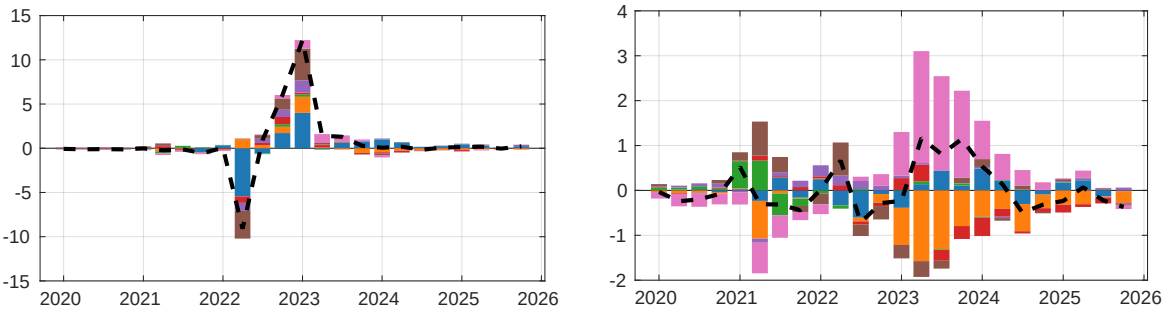
Figure 8: Decomposition of the balance of risk $h = 4$ - PCE



(a) DR (b) QR

Notes: The dashed black line indicates the difference between the predicted balance of risk 4-quarters ahead and as well as the historical average of the 4-quarter ahead balance of risk. The left column presents estimates computed with the density regression (DR) model and the right column results for the quantile regression (QR) model. The stacked bars contain the contributions, with: ● dynamic persistence, ● domestic business cycle, ● global business cycle, ● commodity prices, ● financial conditions, ● monetary policy, ● inflation expectations.

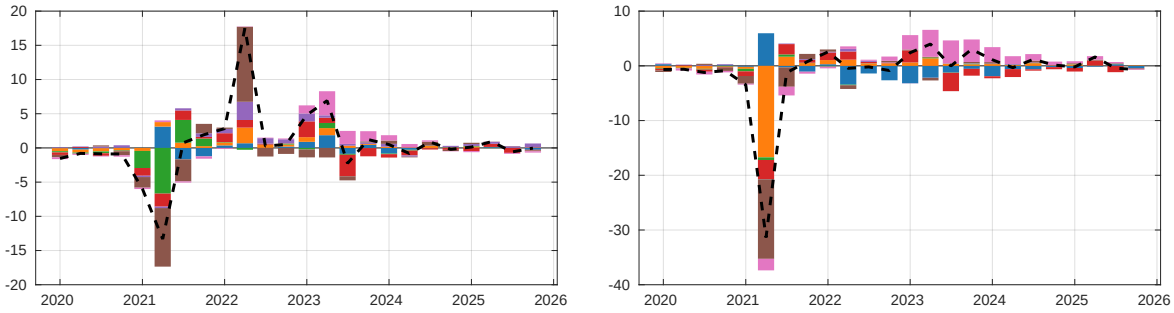
Figure 9: Decomposition of the balance of risk $h = 4$ - CPI excluding energy and food



(a) DR (b) QR

Notes: The dashed black line indicates the difference between the predicted balance of risk 4-quarters ahead and as well as the historical average of the 4-quarter ahead balance of risk. The left column presents estimates computed with the density regression (DR) model and the right column results for the quantile regression (QR) model. The stacked bars contain the contributions, with: ● dynamic persistence, ● domestic business cycle, ● global business cycle, ● commodity prices, ● financial conditions, ● monetary policy, ● inflation expectations.

Figure 10: Decomposition of the balance of risk $h = 4$ - CPI



(a) DR (b) QR

Notes: The dashed black line indicates the difference between the predicted balance of risk 4-quarters ahead and as well as the historical average of the 4-quarter ahead balance of risk. The left column presents estimates computed with the density regression (DR) model and the right column results for the quantile regression (QR) model. The stacked bars contain the contributions, with: ● *dynamic persistence,* ● *domestic business cycle,* ● *global business cycle,* ● *commodity prices,* ● *financial conditions,* ● *monetary policy,* ● *inflation expectations.*

5.4 Alternative Parameter Settings for Risk Measures

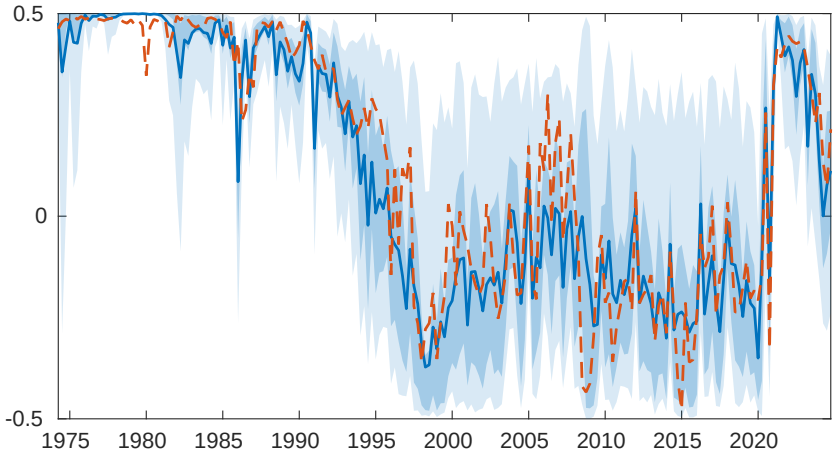
This appendix section presents the balance of risk and its decomposition for alternative parameter settings, with $\alpha = \beta = 0$. Under these settings, the risk measure reflects a purely statistical perspective, corresponding to a central bank that considers only the probability of inflation overshooting or undershooting its target, irrespective of the magnitude of the deviation. From an intuitive perspective, this has important implications as illustrated by the example in Kilian and Manganelli (2008). Specifically, assume that a central banker faces two scenarios: (a) 2.001% inflation with certainty, or (b) 10% inflation with 20% probability and inflation less than 2% with 80% probability. If $\alpha = \beta = 0$, then (a) is considered worse, since $|BR_{0,0}^{(a)}| = |0.5| > |BR_{0,0}^{(b)}| = |-0.3|$. In practice, however, central bankers would likely prefer scenario (a) over (b), implying that $\alpha > 0$ and $\beta > 0$, as considered in the main body of the paper.

Nonetheless, comparing the results obtained under $\alpha = \beta = 0$ with those from, for example, the baseline specification in the main text ($\alpha = \beta = 2$) remains informative, as it highlights the role of risk aversion in shaping the balance of risk. While the $\alpha = \beta = 0$ case captures a purely probabilistic notion of balance, the $\alpha = \beta = 2$ specification gives greater weight to large deviations from target, reflecting the curvature of a standard quadratic loss function. Comparing the two therefore illustrates how introducing policy-relevant risk aversion shifts the interpretation of "balanced risks" from a purely statistical to a welfare-oriented perspective, helps disentangle the relative importance of the frequency versus the magnitude of deviations, and allows an assessment of the robustness of the results to alternative preference specifications. Comparing different parameter settings is thus particularly useful when one does not wish to, or cannot, determine a specific parametrization.

Similarly, the choice of parameter settings plays an important role in defining contributions to risk, as briefly discussed in the main body of the paper. Consider a simple example: if a central bank is only concerned with deflation risks ($w = 1$), but an increase in commodity prices shifts probability mass within the upper tail of the distribution, thereby increasing the likelihood of very high inflation events, then commodity prices should not emerge as a predictor of risk. Furthermore, the relative size of the contributions will vary with the preference parameters and with how individual predictors shift probability mass. For instance, if a variable shifts mass toward extreme inflation realizations, its relative contribution will be larger for higher levels of risk aversion. Alternatively, under the same preference parameters, a predictor that shifts mass further away from the target will have a larger contribution than one

that shifts the same mass closer to the target. Therefore, analogous to the risk measures themselves, the contributions to risk cannot be interpreted without taking into account risk preferences. The requirement that any risk measure must be linked to the preferences of the economic agent, as stated in Machina and Rothschild (1987), thus generalizes to the risk decompositions. Comparing contributions across risk measures can, in turn, yield insights into which variables primarily affect the likelihood versus the magnitude of inflation over- or undershoots, thereby offering a more nuanced understanding of the underlying sources of inflation risk.

Figure 11: Annualized Quarterly Inflation

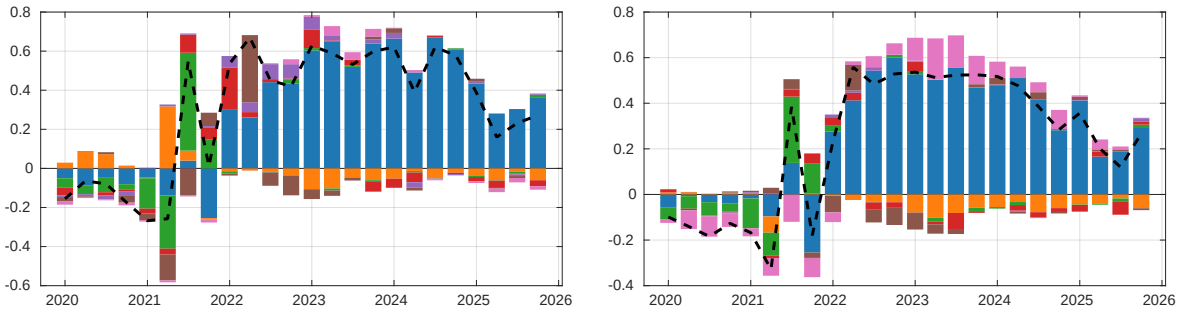


Notes: The solid blue line indicates the estimates for the density regression model over the full sample obtained. The dark shaded area corresponds to the 50% and the light shaded area to the 90% credible interval. The dashed red line shows the corresponding estimates computed with the QR model.

Figure 11 shows the results for the risk measures generated with both the DR and QR models. Overall, the two measures move broadly together. Up to around 1990, both models indicate elevated inflation risks, with a brief decline around 1986. Starting around 1991, i.e., following the end of the first Gulf War, risks gradually normalize before shifting toward deflation risks around 1995. Both models display a temporary low around 1997, roughly coinciding with the Asian financial crisis. Thereafter, risks remain mostly on the downside, interrupted by a series of increases between 2005 and the onset of the global financial crisis, potentially corresponding to the sequence of oil price shocks during that period, before dropping sharply with the financial crisis itself. Until the onset of the pandemic, when risks rise abruptly toward pronounced inflation risks, they remain predominantly on the downside, normalizing again toward the end of the sample. Larger discrepancies between the two measures emerge around 1980, the mid-1990s, and in the

run-up to and during the global financial crisis. Overall, the QR-based risk measure appears somewhat more volatile during the second half of the sample. The observed divergences coincide with structural breaks, abrupt policy shifts, and tail-driven shocks, which may have altered the shape of the inflation distribution and, in turn, affected the behaviour of the quantile-based and distributional models differently.

Figure 12: Decomposition of the balance of risk $h = 4$ and $\alpha = \beta = 0$ - PCE excluding energy and food



(a) DR (b) QR

Notes: The dashed black line indicates the difference between the predicted balance of risk 4-quarters ahead and as well as the historical average of the 4-quarter ahead balance of risk. The left column presents estimates computed with the density regression (DR) model and the right column results for the quantile regression (QR) model. The stacked bars contain the contributions, with: ● *dynamic persistence,* ● *domestic business cycle,* ● *global business cycle,* ● *commodity prices,* ● *financial conditions,* ● *monetary policy,* ● *inflation expectations.*

The corresponding decomposition results for the pandemic sample are shown in Figure 12. Overall, the decompositions are consistent across the DR and QR models and qualitatively align with the results presented in the main body of the paper. Differences in the magnitude of individual contributions, both here and in the main analysis, primarily reflect the specification of the loss function. With $\alpha = \beta = 0$, only the direction in which a variable shifts probability mass relative to the target matters, whereas with $\alpha = \beta = 2$, the magnitude of that shift also influences the risk measure, as discussed previously. Bars that appear relatively larger under $\alpha = \beta = 0$ than under $\alpha = \beta = 2$ therefore reflect mass shifts occurring in closer proximity to the target. For a more detailed analysis, the density decompositions presented in the paper can be computed to provide a complete cross-sectional view.

In sum, the alternative parametrization delivers broadly similar decomposition patterns, reinforcing the robustness of the main findings.

6 Additional Algorithms

6.1 MCMC Sampler

This section provides a MCMC algorithm for the density regression model analogous to the Variational Bayes algorithm used in the paper, following Rigon and Durante (2021).

Algorithm 4: MCMC algorithm

[0] Initialize the model parameters.

begin

[1] Assign each observation $t = 1, \dots, T$ to a mixture component $c = 1, \dots, C$:

for $t = 1 : T$ **do**

Sample $G_t \in 1, \dots, C$ from a categorical distribution with probabilities

$$\text{pr}(G_t = c|-) = \frac{\left[\nu_c(\boldsymbol{x}_t) \prod_{l=1}^{c-1} \{1 - \nu_l(\boldsymbol{x}_t)\}\right] \sqrt{\tau_c} \left[\sqrt{\tau_c} \{y_t - \boldsymbol{x}_t' \boldsymbol{\beta}_c\}\right]}{\sum_{q=1}^{C} \left[\nu_q(\boldsymbol{x}_t) \prod_{l=1}^{q-1} \{1 - \nu_l(\boldsymbol{x}_t)\}\right] \sqrt{\tau_q} \left[\sqrt{\tau_q} \{y_t - \boldsymbol{x}_t' \boldsymbol{\beta}_q\}\right]},$$

for every $c = 1, \dots, C$.

end

[2] Sample the parameters $\boldsymbol{\psi}_c$ and the horseshoe prior parameter $\boldsymbol{\upsilon}_c$, τ_c, $\boldsymbol{\eta}_c$, and ξ_c for each $c = 1, \dots, C-1$:

for $c = 1, \dots, C-1$ **do**

[2a] Sample the parameters $\boldsymbol{\psi}_c$ for $c = 1, \dots, C-1$ exploiting the continuation-ratio parametrization and the results for Bayesian logistic regression in Polson et al. (2013):

for *t such that* $G_t > c - 1$ **do**

Sample the Pólya-gamma distributed data $\omega_{t,c}$ from $(\omega_{t,c}|-) \sim PG(1, \boldsymbol{z}_t' \boldsymbol{\psi}_c)$.

end

Update ψ_c from $(\psi_c|-) \sim N(\boldsymbol{\mu}_{\psi,c}, \boldsymbol{\Sigma}_{\psi,c})$ by Pólya-gamma augmentation, where

$\boldsymbol{\mu}_{\psi,c} = \boldsymbol{\Sigma}_{\psi,c} [\boldsymbol{X}' \boldsymbol{\kappa}_c]$, $\boldsymbol{\Sigma}_{\psi,c} = \left[\boldsymbol{X}' diag(\omega_{1,c}, \dots, \omega_{\bar{T}_c,c}) \boldsymbol{X} + \tilde{\Upsilon}_c\right]^{-1}$, with

$\boldsymbol{\kappa}_c = (\bar{z}_{1,c} - 0.5, \dots, \bar{z}_{\bar{T}_c,c} - 0.5)'$, $\bar{z}_{t,c} = 1$ if $G_t = c$ and $\bar{z}_{t,c} = 0$ if $G_t > c$, and $\tilde{\boldsymbol{\Upsilon}}_c = diag(\bar{\tau}^{-2} \boldsymbol{\upsilon}_c^{-2})$.

[2b] Update the Horseshoe prior parameters following Makalic and Schmidt (2016) from:

$(\upsilon_{j,c}^2|-) \sim IG\left(1, \frac{\psi_{j,c}^2}{2\bar{\tau}_c^2} + \frac{1}{\eta_{j,c}}\right)$, $(\eta_{j,c}|-) \sim IG\left(1, 1 + \frac{1}{\lambda_{j,c}^2}\right)$

$(\bar{\tau}_c^2|-) \sim IG\left(\frac{K+1}{2}, \frac{1}{\xi_c} + \sum_{j=1}^{k} \frac{\psi_{j,c}^2}{2\upsilon_{j,c}^2}\right)$, $(\xi_c|-) \sim IG\left(1, 1 + \frac{1}{\bar{\tau}_c^2}\right)$.

end

[3] Sample the parameters $\boldsymbol{\beta}_c$ from $(\boldsymbol{\beta}_c|-) \sim N(\boldsymbol{\mu}_{\beta_c}, \Sigma_{\boldsymbol{\beta}_c})$, where

$\boldsymbol{\mu}_{\beta_c} = \boldsymbol{\Sigma}_{\beta_c} \left(\tau_c \boldsymbol{X}_c' \boldsymbol{y}_c + \underline{\boldsymbol{\Sigma}}_{\beta}^{-1} \underline{\boldsymbol{\mu}}_{\beta}\right)$, $\boldsymbol{\Sigma}_{\beta_c} = \left(\tau_c \boldsymbol{X}_c' \boldsymbol{X}_c + \underline{\boldsymbol{\Sigma}}_{\beta}^{-1}\right)^{-1}$. $\boldsymbol{X}_c$ and $\boldsymbol{y}_c$ correspond to the observations for which $G_t = c$.

[4] Draw the precision parameter τ_c for $c = 1, \dots, C$:

for $c = 1, \dots, C$ **do**

Sample τ_c from $(\tau|-) \sim G\left(a_\tau + \frac{1}{2} \sum_{t=1}^{T} \mathbb{I}(G_t = c), b_\tau + \frac{1}{2} \sum_{t:G_i=c} \{y_t - \boldsymbol{x}_t' \boldsymbol{\beta}_c\}^2\right)$

end

end

6.2 Computing Quantiles for Density Regression

By the properties of the normal distribution, the sum of multiple normals is again normal. The CDF and PDF are hence available as the sum of the appropriately reweighted individual mixture components, as shown above. However, this convenient property does not hold for the quantile function. An analytical solution is hence not

available, however, quantiles can be computed using simple root finding procedure, where

$$\Phi_x(y_{t+h})\big|_{\hat{Q}_x^\tau(y_{t+h})} - \tau = 0, \tag{6}$$

where τ denotes a percentile of interest and $\Phi_x(y_{t+h})$ denotes the CDF of the mixture model that is evaluated at a candidate solution $\hat{Q}_x^\tau(y_{t+h})$. The desired percentile, $Q_x^\tau(y_{t+h})$, is the solution to Equation (6). Note that because the CDF is proper and the entire distribution is estimated simultaneously, the resulting conditional quantiles do not cross.