Inflation Dynamics Beyond the Mean: A Time-Varying Multiple-Quantile Phillips Curve with Quantile-Heterogeneous Slopes

Tibor Pál

May 2026

Abstract

This paper studies the dynamics of the conditional distribution of US inflation within a hybrid New Keynesian Phillips curve framework. Conditional inflation quantiles are jointly modeled using a dynamic multiple quantile specification. To accommodate time variation and quantile heterogeneity, the smoothed Dynamic Multiple Quantile model is introduced that allows the Phillips curve slope to vary over time and across the distribution. Using quarterly data from 1961Q1 to 2025Q4, we document pronounced nonlinearities in the Phillips curve and in the relative importance of backward and forward-looking inflation expectations. The estimated slopes are asymmetric across quantiles, with a marked flattening in the upper tail. The entire cross-quantile profile evolves over time, implying a time-varying Phillips curve shape. Periodic flattening of the Phillips curve, particularly in the lower tail, is robust across specifications and consistent with the missing disinflation following the Great Recession and in the post-pandemic period. Lastly, the filtered conditional inflation quantiles, together with higher-order moments, provide a quantitative measure of inflation tail risk.

Keywords: Phillips Curve; Inflation Dynamics; Dynamic Multiple Quantiles; Score-driven models; Inflation at risk
JEL Classification: E31, C14, C32, E52

1 Introduction

Price stability is a central objective of monetary policy, and the evaluation of inflation risk is an important input to fulfill this mandate. Understanding the conditional distribution of inflation, its determinants, and their dynamics is therefore crucial for policy decisions.

Since the original contribution of Phillips (1958), particular interest has focused on identifying the key drivers of inflation dynamics, with emphasis on the relative roles of forward- and backward-looking price-setting behavior and the strength of the link between real activity and inflation. Despite substantial methodological progress, empirical findings remain mixed, and a consensus has yet to emerge. This is particularly concerning given the widespread imposition of this theoretical link in studies aiming to identify quantities related to equilibrium or long-term concepts such as the output gap or the natural rate of unemployment.

Given this context, this paper contributes to the literature on inflation dynamics by jointly modeling multiple quantiles of the US inflation within a framework motivated by the hybrid New Keynesian Phillips Curve. Building on the Dynamic Multiple Quantile approach, the model allows the effects of inflation determinants to vary both over time and across quantiles corresponding to predefined probability levels.

Traditionally, financial econometricians have focused on studying the behavior of conditional distributions in the tails, a focus that has become increasingly relevant in macroeconomic contexts. Building on the well-established concept of value at risk (VaR), a risk metric that quantifies the highest potential loss of a financial portfolio or asset within a specific time frame at a chosen probability threshold, research into the tail-end behavior of macroeconomic variables has received growing attention.

At-risk frameworks, such as growth-at-risk (Adrian et al., 2019), inflation-at-risk (López-Salido and Loria, 2024), and debt-at-risk (Furceri et al., 2025), provide valuable tools to assess the conditional distribution of macroeconomic variables. However, these studies typically rely on time-invariant reduced form quantile regressions, in which conditional distributions are linked to a set of explanatory variables. Specifically, they are based on static specifications with parameters that are constant within each estimation period and exhibit limited state dependence.

Consequently, while these models provide insight into how current conditions map into future distributions, they remain limited in their ability to capture the time-varying and state-dependent nature of quantile dynamics. These limitations are especially pronounced when modeling inflation dynamics, where persistence is a well-documented feature implying a gradually evolving conditional inflation distribution through the propagation of past shocks.

Therefore, in contrast to the standard quantile regression approach, and aligned with a time-series setting, we consider a dynamic quantile model in which the quantile of the distribution of the random variable under study is conditioned on the available prior-period information, $p(y_t \mid \mathcal{I}_{t-1})$, where $\mathcal{I}_{t-1}$ contains the entire history of y_t up to time $t-1$, thereby introducing an explicit backward-looking information component.

The proposed model builds on the Dynamic Multiple Quantile (DMQ) framework of Catania and Luati (2023), with two key departures. First, in line with the hybrid New Keynesian Phillips Curve (NKPC), additional variables are included in the specification, namely proxies for marginal cost and inflation expectations. Second, reflecting theoretical considerations underlying the dynamic Phillips curve relationship, the slope parameter varies over time and across quantiles. In addition, inflation inertia and the weight on inflation expectations vary over time, making explicit the conditional role of backward- and forward-looking components.

An inherent feature of the DMQ model is its reliance on a large set of quantiles to provide sufficient information for updating the dynamic quantiles. Introducing quantile-specific parameters in the original framework therefore results in a substantial increase in model dimensionality. This introduces a trade-off between retaining sufficient information to update the dynamic quantiles and maintaining parsimony in the parameter space. In the proposed framework, this trade-off is relaxed by specifying parameters at a selected subset of probability levels and interpolating them at the remaining quantiles. Therefore, the developed DMQ specification allows the parameters associated with inflation determinants to vary over time and across quantiles while keeping the overall parametrization tractable. The model is interpreted as an observation-driven approximation and is labeled as the Smoothed DMQ (sDMQ) model, where *smoothed* refers to the use of monotone interpolation to construct a continuously differentiable parameter curve across probability levels, thereby ensuring smooth parameter variation over the conditional inflation distribution.

The main contributions of the study are as follows. First, we generalize the DMQ framework by allowing parameters to vary jointly across quantiles and over time, yielding a time-varying and heterogeneous DMQ specification. Second, we employ this framework to analyze US inflation dynamics using specifications motivated by the New Keynesian Phillips Curve. Third, we provide empirical evidence of state-dependent inflation inertia, the role of inflation expectations, and nonlinear Phillips curve dynamics across both time and the conditional distribution, including a time-varying cross-quantile shape and a persistently flat slope in the upper tail. Fourth, we demonstrate that the choice of expectation proxy is crucial for identification: short-run expectations generate richer higher-order dynamics, whereas more anchored long-run expectations attenuate tail behavior.

A key challenge in estimating the New Keynesian Phillips curve using limited-information econometric methods is the limited identifying power of economic aggregates, as emphasized by Mavroeidis et al. (2014). This paper shows that incorporating distributional information and time variation can provide additional insights into inflation dynamics beyond conventional aggregate approaches.

Compared to a traditional Phillips curve model, the proposed framework offers several advantages. First, it captures time-varying and heterogeneous responses of inflation to its determinants. For example, the Phillips curve relationship at upper quantiles can differ markedly from that at lower quantiles, which is relevant for monetary policy. Second, the model characterizes the dynamics of the conditional inflation distribution, providing information beyond the first two moments that is useful for monetary policy and risk management.

The empirical analysis, based on quarterly US core inflation, reveals clear heterogeneity in Phillips curve relationships across quantiles. The results indicate an asymmetric Phillips curve that flattens in the upper tail and exhibits time variation concentrated in the lower and middle parts of the distribution. This profile evolves over time, with both its shape and magnitude shifting across macroeconomic regimes. Using alternative expectation measures, the short-run specification yields less persistent slope estimates, while the long-run specification delivers slopes with greater and more persistent variability. The overall slope profile and its time variation remain robust across specifications.

During the 1970s, an increase in the weight on inflation expectations is consistent with weakly anchored expectations and the high-inflation regime of the Great Inflation. Following the Global Financial Crisis, the flattening of the Phillips curve in the lower tail, combined with strong inflation persistence, is consistent with the *missing disinflation* observed during the recovery. In the post-COVID period, further changes in the slope profile and a renewed role for expectations highlight the importance of state dependence and time variation in inflation dynamics.

2 The econometric framework

The econometric framework for modeling inflation quantiles builds on the Dynamic Multiple Quantile (DMQ) model of Catania and Luati (2023). A key limitation of the DMQ model is its reliance on a large set of quantiles to update the time-varying quantile processes. When the focus is on studying dynamic relationships with semi-structural heterogeneity across the distribution, increasing the size of the jointly modeled quantile set reduces the flexibility of the model specification.

To address this shortcoming, the proposed model imposes a more parsimonious parametrization. Specifically, parameters of the explanatory variables are specified only at selected quantiles corresponding to predetermined probability levels and are subsequently interpolated for the remaining quantiles. Focusing on a reduced set of quantile processes facilitates the analysis of relationships across the distribution and opens the way to introduce time-varying parameters.

The section begins with a brief overview of the dynamic quantile Phillips curve model before outlining its main building blocks in the following subsections. The baseline DMQ model, its properties, and its limitations are then discussed in detail, motivating the subsequent modifications that lead to the proposed sDMQ model.

Lastly, the sDMQ model with time-varying parameters is presented, providing an effective framework for analyzing the dynamics of inflation quantiles. The methodology incorporates additional variables in line with a New Keynesian Phillips curve relationship, with the slope parameter varying across quantiles and over time.

2.1 The dynamic multiple quantile Phillips Curve model

This section introduces the general structure for modeling the dynamics of the inflation distribution within a DMQ framework. The model is consistent with a hybrid New Keynesian Phillips Curve (NKPC), relating inflation to both forward- and backward-looking expectations and allowing for a trade-off between the output gap and inflation.

The main departure from the baseline DMQ specification is the inclusion of exogenous variables whose effects on the modeled quantiles are allowed to be heterogeneous and time varying. This objective introduces a key difficulty. The original DMQ model requires joint modeling of a large set of quantiles to obtain sufficient information for the score-driven updates, thereby limiting the feasible parameter dimensionality.

The proposed approach addresses this econometric challenge by focusing on a subset of quantiles at selected locations of the conditional distribution. Throughout the remainder of the paper, these quantiles of interest are referred to as *target quantiles*, and the parameters governing the structural relationships at the associated probability levels are referred to as *target parameters*. The parameter values at the remaining quantiles are obtained by interpolation, implying that these *auxiliary quantiles* are governed by *auxiliary parameters* that are deterministic, smooth functions of the target parameters. This construction restores a large set of quantiles, providing sufficient information for updating the target quantiles. As a result, the sDMQ specification enables the analysis of heterogeneous and dynamic responses of inflation to its determinants across different parts of the inflation distribution.

Quantiles divide the conditional distribution into regions such that a given proportion of observations falls below each quantile value at a specific probability level. In other words, the p-th conditional target inflation quantile $\hat{q}^{\tau_p}(\pi)$ is the inflation rate at which the conditional probability of inflation being less than $\hat{q}^{\tau_p}(\pi)$ equals τ_p. Meanwhile, the probability of inflation being greater than $\hat{q}^{\tau_p}(\pi)$ is equal to $1 - \tau_p$.

Consequently, a general NKPC-based specification for the conditional quantile function of inflation, $\hat{q}_t^{\tau_p}(\pi)$ is given by

$$\hat{q}_t^{\tau_p}(\pi) = \tilde{q}_t^{\tau_p}(\pi) + \kappa_{p,t}\, x_t + \delta_t\, E_t(\pi_{t+1}), \tag{2.1}$$

where $\hat{q}_t{}^{\tau_p}(\pi)$ denotes the pth conditional target inflation quantile at the probability level τ_p, given the information set available at time $t-1$. The term x_t represents a proxy for real marginal cost, with a time-varying slope parameter $\kappa_{p,t}$ specific to each quantile process. The variable $E_t(\pi_{t+1})$ denotes the forward-looking inflation expectations with a time-varying weight δ_p. The term $\tilde{q}_t^{\tau_p}(\pi)$ denotes the score-driven *dynamic quantiles*, each with its own updating process, described in the following section.

To conclude the section and for ease of reference, Table 1 summarizes definitions of the key quantile objects and parameters employed in the sDMQ framework.

Table 1: Notation and Definitions

Term	Notation	Definition
Target quantiles	$\hat{q}_t^{\tau_p}$	Directly modeled semi-structural conditional quantiles.
Auxiliary quantiles	$\bar{q}_t^{\tau_j}$	Quantiles not directly parametrized.
Dynamic quantiles	$\tilde{q}_t^{\tau_j}$	Score-driven quantiles driving the target dynamics.
Target parameters	$\hat{\kappa}_{p,t}$	Time-varying slope parameters at target quantiles.
Auxiliary parameters	$\bar{\kappa}_{j,t}$	Time-varying slope parameters at auxiliary quantiles.

2.2 The score-driven update of the DMQ model

Several features make the DMQ framework attractive for jointly studying the dynamics of the inflation quantiles. First, the model satisfies the non-crossing quantile criterion by construction. Second, it is consistent with the defining property of fixed quantiles, ensuring that in the limit the proportion of observations falling below the unconditional τ-level quantile converges to τ. Third, the model can handle a large set of quantiles simultaneously, ensuring that it incorporates sufficient information to drive the evolution of the time-varying quantiles.

Yet the main novelty of the DMQ model lies in its use of a score-driven updating scheme. Formally, Catania and Luati (2023) define the updating process as

$$\tilde{q}_t^{\tau_j} = \begin{cases} \tilde{q}_t^{\tau_{j+1}} - \eta_{j,t}, & \text{if } \tau_j < \tau_{j^*} \\ \tilde{q}_t^{\tau_{j^*}}, & \text{if } \tau_j = \tau_{j^*} \\ \tilde{q}_t^{\tau_{j-1}} + \eta_{j,t}, & \text{if } \tau_j > \tau_{j^*} \end{cases} \tag{2.2}$$

with

$$\tilde{q}_t^{\tau_{j^*}} = \tilde{q}_0^{\tau_{j^*}}(1-\beta) + \alpha u_{t-1}^{\tau_{j^*}} + \beta \tilde{q}_{t-1}^{\tau_{j^*}} \tag{2.3}$$

where $\tilde{q}_0^{\tau_{j^*}}$ denotes the unconditional quantile at the reference probability level, β is the inertia parameter, and α controls the weight of the forcing variable. The process serves as the *reference quantile*, defining the dynamics of the remaining quantiles through

$$\eta_{j,t} = \exp(\xi_{j,t}) \tag{2.4}$$

which are positive-valued random variables with

$$\xi_{j,t} = \xi_{j,0}(1-\phi) + \gamma u_{t-1}^{\tau_j} + \phi \xi_{j,t-1} \tag{2.5}$$

where ϕ and γ denote the inertia and updating weight parameters, respectively.

The quantile-specific terms $u_t^{\tau_{j^*}}$ and $u_t^{\tau_j}$ are the *forcing variables* responsible for the filtering update defined as

$$u_t^{\tau_j} = \begin{cases} b_j a_j^{-1} \sum_{i=1}^{j} z_{i,t}, & \text{if } j < j^* \\ a_j^{-1} \sum_{i=1}^{j} z_{i,t}, & \text{if } j = j^* \\ b_j a_j^{-1} \sum_{i=j}^{j} z_{i,t}, & \text{if } j > j^* \end{cases} \tag{2.6}$$

where $b_s = \mathbb{1}(\tau_j < \tau_{j^*}) - \mathbb{1}(\tau_j > \tau_{j^*})$ and a_j is a normalizing term defined as in Catania and Luati (2023). This normalization is analogous in spirit to inverse information scaling in score-driven models. However, it is based on the unconditional second moments of empirical quantiles rather than on a parametric likelihood.

The *hit variable* $z_{i,t}$ takes the form of

$$z_t = \frac{\partial}{\partial q_t} \sum_{j=1}^{J} \rho_{\tau_j}\left(y_t - q_t^{\tau_j}\right) = \sum_{j=1}^{J} \frac{\partial}{\partial {q_t}^{\tau_j}} \rho_{\tau_j}(y_t - {q_t}^{\tau_j}) = \mathbb{1}(y_t \leq q_t^{\tau_j}) - \tau_j \tag{2.7}$$

where $z_t = \left(z_{i,t},\ i = 1, \ldots, J\right)'$ and $\rho_\tau(\chi) = \chi\left(\tau - \mathbb{1}(\chi < 0)\right)$ is the quantile check function with the indicator function $\mathbb{1}(\cdot)$.

Importantly, the sequences of the hit variable $\{z_t\}_{t\in\mathbb{N}}$ and the corresponding forcing variable $\{u_t^{\tau_j}\}_{t\in\mathbb{N}}$ are independent and identically distributed (iid) with zero mean and constant variance. Consequently, the updating mechanism of the DMQ model is consistent with the general class of observation-driven models (Creal et al., 2013; Harvey, 2013). Hence, the model provides an observation-driven representation of a predefined set of quantiles within the score-driven framework.

2.3 Inflation inertia and expectations dynamics

To introduce state dependence in the NKPC relationships, first we allow the inertia parameter β, associated with the dynamics of the reference quantile, to vary over time. In addition, we reassign the forward-looking component in equation 2.1 by replacing the unconditional quantile at the reference probability level $\tilde{q}_0^{\tau_{j^*}}$ in equation 2.3 with inflation expectations $E_t(\pi_{t+1})$. This yields a standard and economically coherent identification strategy without expanding the parameter space, while providing a convenient way to quantify the relative weight of inflation inertia and forward-looking inflation expectations.

Accordingly, the reference quantile process in equation 2.3 is redefined as

$$\tilde{q}_t^{\tau_{j^*}} = \left(1 - \exp(\beta_t)\right) E_t(\pi_{t+1}) + \alpha u_{t-1}^{\tau_{j^*}} + \exp(\beta_t) \tilde{q}_{t-1}^{\tau_{j^*}} \tag{2.8}$$

and the updating equation for the time varying weight parameter takes the form of

$$\beta_t = (1 - \phi_\beta)\beta_0 + \phi_\beta \beta_{t-1} + \rho_\beta u_t^\beta \tag{2.9}$$

where ϕ_β denotes the inertia parameter, ρ_β the loading coefficient on the forcing variable u_t^β, and β_0 the long-run reference weight, capturing the relative importance of backward- and forward-looking components in conditional inflation dynamics.

Lastly, the innovation u_t^β is defined as

$$u_t^\beta = \frac{\partial}{\partial \beta_t} \sum_{j=1}^{J} \rho_{\tau_j}\big(y_t - q_t^{\tau_j}\big) = \sum_{j=1}^{J} \frac{\partial \rho_{\tau_j}}{\partial q_t^{\tau_{j^*}}} \frac{\partial q_t^{\tau_{j^*}}}{\partial \beta_t} = z_t \exp(\beta_t)\big(\tilde{q}_{t-1}^{\tau_{j^*}} - E_t(\pi_{t+1})\big). \tag{2.10}$$

It should be noted that, by construction, beyond its direct effect on the reference quantile, inflation expectations affect the dynamics of the remaining quantile processes indirectly through the reference quantile, in a manner analogous to the inertia parameter β in the baseline DMQ specification.

2.4 Quantile-heterogeneous and time-varying slopes

The stability of the slope of the Phillips curve remains a topic of ongoing debate. Although some studies find little evidence of significant changes in recent decades (e.g., Hazell et al., 2022), others point to structural changes within the New Keynesian framework, including changes in the Phillips curve itself. For example, Ball and Mazumder (2011) argue that inflation dynamics changed notably after the global financial crisis (GFC), and Rossi et al. (2024) provides evidence of a marked decline in the slope of the Phillips curve since the 1980s. Given these conflicting findings, the question of whether the Phillips curve has genuinely flattened over time remains unresolved.

A natural way to address these considerations within the present framework is to allow the Phillips curve slope parameter to vary across time and quantiles. However, as shown, the score-driven update of time-varying processes requires modeling a large set of quantiles. This poses a nontrivial challenge to the objectives of the study because the original DMQ specification would require a substantial expansion of the parameter space, thereby creating a dimensionality problem. Assigning common parameters to broader quantile ranges mitigates this burden. However, it induces substantial bias by imposing a non-smooth structure on the slope parameters across the distribution, thereby limiting their economic plausibility.

To address this challenge, an alternative approach is adopted to reduce the number of required parameters while retaining sufficient heterogeneity in the slope parameters and enhancing smoothness across quantiles. Specifically, a reduced set of parameter processes, $\hat{\kappa}_{p,t}$, is specified and interpolated to obtain the complete set of quantile-specific parameters at each time t. These time-varying slope parameters are associated with a set of target quantiles defined at predefined probability levels, hereafter referred to as *target probability levels*. Specifically, the analysis considers the standard set $\tau \in \{0.05, 0.25, 0.75, 0.95\}$, which is commonly used in studies of tail risk of macroeconomic variables (e.g., Adrian et al., 2019; López-Salido and Loria, 2024). Additionally, we include the central quantile $\tau = 0.50$, as the median is the standard reference and improves economic interpretation. Empirical evidence in Catania and Luati (2023) further supports this choice, showing that probability levels close to the median minimize the objective function.

The selection of the predefined probability levels serves two primary purposes. First, it ensures representation of the key regions of the distribution while maintaining a balance between parsimony and flexibility. Second, it facilitates identification consistent with theoretical considerations. Under the minimal and economically reasonable assumption that the parameters are monotone between these probability levels, this specification provides sufficient structure to support the interpolation step.

Building on the proposed modifications to the modeling framework, Eq. 2.1 is reformulated using the updated notation as

$$q_t^{\tau_j}(\pi) = \tilde{q}_t^{\tau_j}(\pi) + \exp(\kappa_{j,t})\, x_t, \tag{2.11}$$

where the full set of conditional inflation quantiles $q_t^{\tau_j}(\pi)$ comprises the directly modeled target quantiles $\hat{q}_t^{\tau_p}(\pi)$ at the target probability levels and auxiliary quantiles $\bar{q}_t^{\tau_j}(\pi)$ at the remaining probability levels. Correspondingly, the full set of slope parameters $\kappa_{j,t} = (\kappa_{j,t},\ j = 1, \ldots, J)'$ is constructed from the vector of target parameters $\hat{\kappa}_{p,t} = (\hat{\kappa}_{p,t},\ p = 1, \ldots, P)'$, with the remaining elements obtained by interpolation. Specifically, for non-target probability levels, $\kappa_{j,t}$ is deterministically generated from the target parameters $\hat{\kappa}_{p,t}$, which in turn yields the $J - P$ auxiliary quantiles whose slope parameters are not directly estimated.

Formally,

$$\kappa_{j,t} = \mathbb{1}(\tau_j = \tau_p)\, \hat{\kappa}_{p,t} + \mathbb{1}(\tau_j \neq \tau_p)\, \bar{\kappa}_{j,t}, \tag{2.12}$$

where $\bar{\kappa}_{j,t}$ are obtained by monotone piecewise cubic interpolation (MPCI), as described in Fritsch and Carlson (1980), within the range spanned by the target parameters, and by extrapolation for probability levels below $\tau = 0.05$ and above $\tau = 0.95$. Using this construction, the interpolated parameters are given by

$$\bar{\kappa}_{j,t} = h_{00}(b_j)\, \kappa_t^{\tau_L} + h_{01}(b_j)\, \kappa_t^{\tau_U}, \qquad b_j = \frac{\tau_j - \tau_{p_L(j)}}{\tau_{p_U(j)} - \tau_{p_L(j)}}, \tag{2.13}$$

where $p_L(j) = \arg\max_{p:\, \tau_p < \tau_j} \tau_p$ and $p_U(j) = \arg\min_{p:\, \tau_p > \tau_j} \tau_p$, and the MPCI basis functions are

$$h_{00}(b) = 2b^3 - 3b^2 + 1, \qquad h_{01}(b) = -2b^3 + 3b^2. \tag{2.14}$$

Turning to the specification of the target parameters, their time-varying process follows an autoregressive (AR) process and is defined as

$$\hat{\kappa}_{p,t} = \hat{\kappa}_{p,0} + \phi_p^{\kappa} \hat{\kappa}_{p,t-1} + \rho_p u_{p,t}^{\kappa}, \tag{2.15}$$

where $\hat{\kappa}_{p,0}$ denotes the quantile-specific intercept terms, ϕ_p^{κ} are the inertia parameters, and $u_{p,t}^{\kappa}$ are the forcing variables weighted by ρ_p.

Lastly, the forcing variables are specified as

$$u_{p,t}^{\kappa} = \frac{\partial}{\partial \hat{\kappa}_{p,t}} \sum_{j=1}^{J} \rho_{\tau_j} \left(\pi_t - q_t^{\tau_j} \right) = \sum_{j=1}^{J} \frac{\partial \rho_{\tau_j}}{\partial q_t^{\tau_j}} \frac{\partial q_t^{\tau_j}}{\partial \kappa_{j,t}} \frac{\partial \kappa_{j,t}}{\partial \hat{\kappa}_{p,t}}, \tag{2.16}$$

where the individual components of the forcing variable are obtained as

$$\frac{\partial \rho_{\tau_j}}{\partial q_t^{\tau_j}} = \mathbb{1}(\pi_t \leq q_t^{\tau_j}) - \tau_j = z_{j,t},$$

$$\frac{\partial q_t^{\tau_j}}{\partial \kappa_{j,t}} = \exp(\kappa_{j,t})\, x_t,$$

$$\frac{\partial \kappa_{j,t}}{\partial \hat{\kappa}_{p,t}} = \begin{cases} 1, & \tau_j = \tau_p, \\ h_{00}(b_j)\, \mathbb{1}\{p = p_L(j)\} + h_{01}(b_j)\, \mathbb{1}\{p = p_U(j)\}, & \tau_j \neq \tau_p. \end{cases}$$

The resulting forcing variable is therefore

$$u_{p,t}^{\kappa} = \sum_{j=1}^{J} z_{j,t} \exp(\kappa_{j,t})\, x_t \left[h_{00}(b_j)\, \mathbb{1}\{p = p_L(j)\} + h_{01}(b_j)\, \mathbb{1}\{p = p_U(j)\} \right]. \tag{2.17}$$

At this point, it is important to note that constructing the forcing variable in this manner implies that each target TVP depends on the corresponding interpolated parameters. While this design choice confines the information channel to selected regions of the distribution, it also provides a tractable and transparent way to extract information that feeds into the score-driven updates.

Finally, recall that the semi-structural inflation quantiles and their associated parameters at the target probability levels are of primary interest and are modeled directly. The interpolated parameters primarily serve to generate a sufficiently dense set of auxiliary quantiles that, via the forcing variables, provide the information required to update the directly specified target quantiles. Accordingly, the proposed specification should be interpreted as an observation-driven approximation to the evolving conditional inflation distribution rather than as the exact data generating process. The smoothing structure imposed across quantiles therefore serves primarily as a regularization device that facilitates identification and stable estimation. By confining economic interpretation to the target quantiles and using auxiliary quantiles exclusively within the updating scheme, this construction also helps mitigate potential bias arising from the imposed cross-quantile restrictions.

2.5 Estimation

The estimation of the model uses the theoretical results of Catania and Luati (2023) for the standard DMQ model. This approach adopts the estimation method first proposed by White et al. (2015).

Specifically, the estimation relies on the quantile check loss function. This function can be interpreted as the negative log-likelihood of an asymmetric Laplace–type density (Koenker and Machado, 1999; Poiraud-Casanova and Thomas-Agnan, 2000; Kotz et al., 2001). More generally, it can be viewed as a member of the tick-exponential family introduced by Komunjer (2005). In the absence of an explicit conditional distributional assumption, minimization of the check loss is equivalent to quasi–maximum likelihood (QML) estimation. Thus, the updating scheme can be viewed as a quasi score-driven (QSD) model, which generalizes the standard score-driven framework outlined in Blasques et al. (2023).

The loss function measures the difference between the observed and predicted values at the probability levels j. Accordingly, the static parameters of the model are estimated by minimizing the objective function of

$$\hat{\theta}_T = \arg\min_{\theta\in\Theta} \sum_{t=1}^{T} \sum_{j=1}^{J} \rho_{\tau_j}\left(y_t - q_t^{\tau_j}(\theta)\right), \tag{2.18}$$

where $q_t^{\tau_j}(\theta)$ denotes the jth conditional inflation quantile, encompassing both target and auxiliary quantile processes. The parameter vector θ collects the static parameters governing the updating dynamics of the J score-driven quantiles and the P time-varying target parameters. The objective function evaluates θ to obtain the parameter values that best fit the observed data.

3 Empirical analysis

3.1 Identification and model specifications

Building on the framework presented in Section 2, we consider the full sDMQ NKPC model

$$q_t^{\tau_j}(\pi) = \tilde{q}_t^{\tau_j}(\pi) + \exp(\kappa_{j,t})\, x_{t-1}, \tag{3.1}$$

where $\tilde{q}_t^{\tau_j}(\pi)$ denotes the dynamic quantile component defined in Eqs. 2.2–2.7 with the modification for the reference quantile process introduced in Eqs. 2.8–2.10. The updating scheme for the time-varying Phillips curve slope $\kappa_{j,t}$ is described in Eqs. 2.12–2.17. The marginal cost x_{t-1} is approximated by the output gap. Note that, in the empirical specification, both the output gap and inflation expectations are treated as predetermined and enter the model in lagged form to mitigate contemporaneous endogeneity concerns.

As is well known, US inflation exhibits high persistence. This makes the inertia component a key determinant in the present setting. Specifically, the *intrinsic* *inflation persistence* in the inflation quantile process arises from the AR dependence on past quantiles, $\tilde{q}_{t-1}^{\tau_j*}$ and $\xi_{t-1}^{\tau_j}$ appearing in Eq. 2.3 and 2.5, respectively. Importantly, by conditioning on information up to time $t-1$, these AR terms substantially reduce the risk of misspecification often arising from omitted lagged inflation dynamics.

Beyond estimating the fully specified sDMQ-TV model defined in Eq. 3.1, the empirical analysis considers a sequence of restricted specifications starting from the baseline DMQ model, which abstracts from structural relationships. In addition to enhancing estimation stability and enabling the evaluation of incremental model improvements, this approach also helps improve identification. For clarity, Table 2 summarizes the model specifications considered in this section and their shorthand notation.

Table 2: Summary of model specifications

Model	Description
DMQ	Baseline DMQ model with the specification of Catania and Luati (2023)
DMQ(S/L)	DMQ NKPC with static quantile-homogeneous parameters (β, κ)
DMQ-TV(S/L)	DMQ model with time-varying quantile-homogeneous parameters (β_t, κ_t)
sDMQ(S/L)	sDMQ NKPC with static quantile-heterogeneous slopes (β, κ_p)
sDMQ-TV(S/L)	sDMQ NKPC with time-varying quantile-heterogeneous slopes (β_t, $\kappa_{p,t}$)

Note: S and L denote models with short-run and long-run inflation expectations, respectively.

Augmenting the baseline DMQ model with time-invariant inflation determinants represents a useful intermediate step. First, it provides a basis for evaluating the relevance of moving from a fully reduced-form inflation dynamics toward a more structural specification. Second, the time-invariant estimation of the relative weights of inflation inertia and forward-looking expectations β and the PC slope κ provides a useful anchor for the time-varying specifications (DMQ-TV). Specifically, the parameterization of the time-varying evolution of these parameters, as in Eq. 2.8, allows these estimates to serve as long-run reference values, thereby improving identification. Similarly, by capturing quantile heterogeneity in the Phillips curve slope, the time-invariant sDMQ model provides a natural basis for treating these estimates as long-run reference values in the sDMQ-TV specification.

While the DMQ-TV specification helps assess time variation, the sDMQ model with static parameters allows for the assessment of quantile heterogeneity. Comparing the estimation results of these specifications allows the two dimensions to be assessed in isolation, whereas the sDMQ-TV model enables their joint assessment.

Higher-order moments of the conditional inflation distribution are of particular interest in the present study. To adequately capture distributional shape and tail behavior, $J = 99$ conditional inflation quantiles are considered across all specifications. Among these, $P = 5$ target quantiles at probability levels $\tau_p \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$ are directly estimated in the sDMQ specifications (sDMQ and sDMQ-TV). The remaining $J - P = 94$ quantiles are treated as auxiliary and primarily serve to inform the score-driven updating mechanism.

Consequently, the full set of dynamic conditional inflation quantiles, comprising both target and auxiliary quantiles, is defined as follows:

$$q_t^{\tau_j}(\pi) = \inf\{\, \pi : F_{t|t-1}(\pi) \geq \tau_j \,\}, \quad \tau_j \in \{0.01,\, 0.02,\, \ldots,\, 0.99\}, \quad j = 1, \ldots, J,$$

$$\hat{q}_t^{\tau_p}(\pi) = \inf\{\, \pi : F_{t|t-1}(\pi) \geq \tau_p \,\}, \quad p = 1, \ldots, P,$$

$$\bar{q}_t^{\tau_j}(\pi) = \inf\{\, \pi : F_{t|t-1}(\pi) \geq \tau_j \,\}, \quad \tau_j \in \{0.01,\, 0.02,\, \ldots,\, 0.99\} \setminus \{\tau_p\}.$$

Several restrictions are imposed to strengthen identification and ensure stable parameter estimation in the sDMQ-TV model. First, the coefficients governing the score-driven updating of the Phillips curve slope parameter, namely the weight on the forcing variable, ρ_p, and the AR coefficient, ϕ_p^κ, are restricted to be common across the five target quantiles. Second, the initial values and long-run reference values are fixed to the corresponding time-invariant estimates from the sDMQ model.

Taking these constrains into account, and after reparameterizing Eq. 2.15, the time-varying Phillips curve slope update is given by

$$\hat{\kappa}_{p,t} = (1 - \phi_\kappa)\hat{\kappa}_{p,0} + \phi_\kappa \hat{\kappa}_{p,t-1} + \rho u_{p,t}^\kappa. \tag{3.2}$$

This reparameterization reduces model dimensionality and enhances estimation stability. Importantly, it provides an economically justified approach to setting the initial values of the time-varying parameters, thereby further reducing the effective parameter space. It also improves economic interpretability. When $\phi_\kappa < 1$, $\hat{\kappa}_{p,0}$ represents not merely an initial condition but a long-run reference level, interpretable as the quantile-specific long-run Phillips curve slope. Consequently, the parameters $\hat{\kappa}_{p,0}$ are estimated using the static sDMQ specification and employed in the time-varying model, reinforcing identification.

Given the relative sparsity of inflation realizations in the tails and the numerical instability induced by interpolation, which can exacerbate the non-smoothness of the likelihood function in the DMQ framework, the static slope parameters may be weakly identified. We address this issue along several dimensions. First, rather than relying on extrapolation beyond the extreme probability levels, we fix the slope in the tail regions at the value estimated at the corresponding boundary quantiles. Second, we impose a lower bound on the slope parameters, $\kappa_p \geq 0.0025$, to enhance numerical stability.

Finally, instead of fixing the initial values $\hat{\kappa}_{p,0}$ to the time-invariant sDMQ estimates, we adopt a shrinkage-based parameterization that allows for partial pooling across quantiles. Specifically, $\hat{\kappa}_{p,0}$ is modeled as a convex combination of the quantile-specific estimates $\hat{\kappa}_p$ obtained by the sDMQ specification and a homogeneous long-run reference $\bar{\kappa}$, defined as the cross-quantile average of the five target quantile estimates, with the shrinkage intensity governed by the parameter ψ_s.

Formally,

$$\hat{\kappa}_{p,0} = \bar{\kappa} + \psi_s \left(\hat{\kappa}_p - \bar{\kappa}\right) \tag{3.3}$$

where $\psi_s \in [0, 1]$ determines the degree of slope heterogeneity across quantiles: the closer its value is to unity, the closer the heterogeneity is to that estimated in the static sDMQ specification.

An important aspect of the DMQ framework is the normalization of the forcing variables across quantile processes. Because the construction of the forcing variables $u^{\kappa}_{p,t}$ defined in Eq. 2.16 relies on multiple components, scaling by a single normalizing term a_p^{-1}, as in Eq. 2.6, is no longer appropriate, as it fails to account for the parameter dependence of the forcing variables.

To address this problem, we adopt a data-driven normalization strategy that accounts for the dependence of the variance of the forcing variables on the shrinkage parameter. Specifically, we precompute a variance surface $\hat{V}_p(\psi_s; \hat{\kappa}_p)$ over a grid $\psi_s \in \{0, 0.1, \ldots, 1\}$, where the variance is evaluated under a fixed-parameter specification, i.e., $\rho = 0$. During estimation, this surface is evaluated at the current value of ψ_s via shape-preserving spline interpolation, yielding $\bar{V}_p(\psi_s; \hat{\kappa}_p) = \mathcal{P}(\psi_s; \hat{V}_p)$. The forcing variables are then normalized as $\bar{u}^{\kappa}_{p,t} = \bar{V}_p(\psi_s; \hat{\kappa}_p)^{-1/2} u^{\kappa}_{p,t}$. This procedure ensures that the scaling of the forcing variables appropriately reflects the underlying parameterization, leading to comparable variances across quantiles, which stabilizes the dynamics and improves optimization behaviour.

Taken together, the fully specified sDMQ-TV model is estimated by jointly estimating the following parameter vector, with β_0 and $\hat{\kappa}_{p,0}$ obtained from the sDMQ model and ϕ_β from the DMQ-TV model:

$$\theta = (\alpha, \gamma, \phi, \phi_\kappa, \rho_\kappa, \rho_\beta, \psi_s),$$

where all parameters are constrained to be positive, and α, ϕ, and ϕ_κ are further restricted to lie within the unit interval.

3.2 Data description

To analyze the dynamics of US inflation, the seasonally adjusted annualized quarterly growth rate of the core Personal Consumption Expenditures (PCE) price index is used over the period 1961Q1 to 2025Q4. Core PCE inflation serves as the key measure of inflation for the Federal Reserve (Fed), capturing underlying price trends by excluding volatile food and energy components and providing a comprehensive measure of consumer prices that reflects evolving spending patterns.

Across all specifications, the output gap is defined as the difference between the logarithm of real GDP and the logarithm of potential output, which serves as a proxy for real economic slack. The output gap series is based on estimates from the Congressional Budget Office (CBO). Our choice of the CBO series is motivated by its structural nature, as potential output is derived from a production function framework that explicitly accounts for trend labor input, capital services, and total factor productivity. In addition, the CBO output gap is widely used in empirical macroeconomic analysis, ensuring comparability with the existing literature.

We proxy inflation expectations using the 2-year inflation expectations series from the Federal Reserve Bank of Cleveland, obtained from FRED at a quarterly frequency and available from 1982Q2 onward. Following the approach of Benigno and Eggertsson (2023), the series is extended backward to 1961Q1 using 12-month inflation expectations from the Livingston Survey. As the latter is available at a semiannual frequency, missing quarterly observations are interpolated using a spline-based, curve-preserving method.

In addition, we use long-run PCE inflation expectations as an alternative longer-horizon measure of expected inflation to assess the sensitivity of the results to the expectations specification and facilitate comparison across alternative expectation measures. Following the approach of Clark and Doh (2014), also adopted by Negro et al. (2017), the series is constructed by combining Survey of Professional Forecasters (SPF) data from 2007Q1 onward with survey-based long-run (5- to 10-year-ahead) PCE inflation expectations from the Federal Reserve Board's FRB/US econometric model for earlier periods.

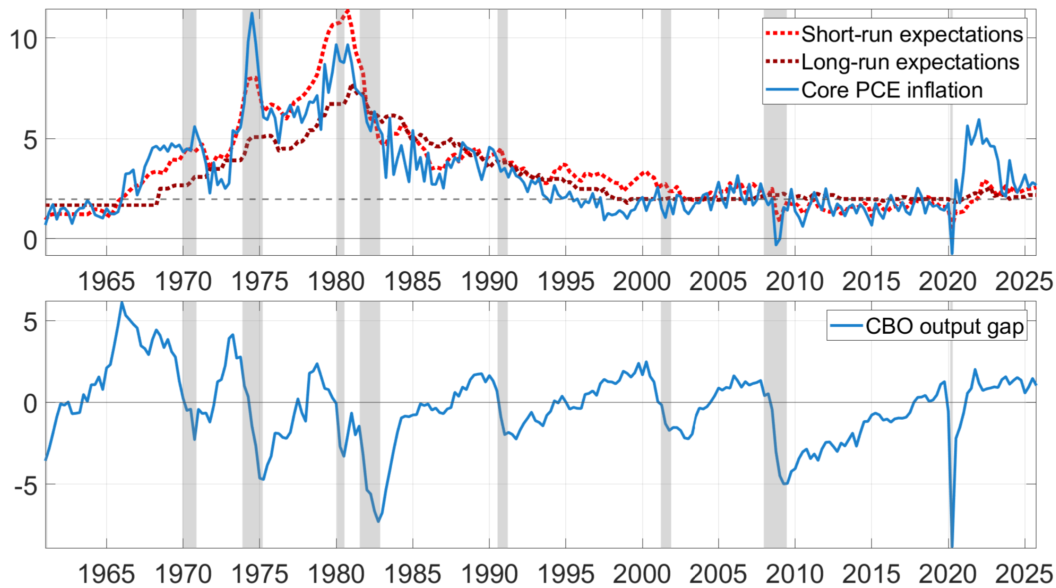

Figure 1: US core PCE inflation and inflation expectations (upper panel) and CBO output gap (lower panel), 1961Q1–2025Q4.

Figure 1 displays the inflation series and Phillips curve determinants used in the analysis. The upper panel shows US core PCE inflation alongside short-run and long-run inflation expectations over the sample period. Core PCE inflation is highly volatile during the 1970s and early 1980s, followed by a sustained moderation from the mid-1980s, and a renewed increase in the post-COVID period. Long-run inflation expectations co-move with PCE inflation but exhibit smoother dynamics and attenuated short-run fluctuations. In contrast, short-run expectations display greater variability and respond more strongly to cyclical movements in inflation.

The lower panel reports the CBO output gap over the sample period. The series exhibits pronounced cyclical dynamics, with large negative gaps during major recessions, notably in the early 1980s, the Great Recession, and the COVID-19 shock in 2020. Overall, the figure highlights the joint evolution of inflation, expectations, and real economic slack, which motivates the empirical Phillips curve specification.

Table 3: Summary statistics of inflation and its determinants (1961Q1–2025Q4)

Variable	$\bar{\mu}$	$\bar{\sigma}$	$\bar{\zeta}$	$\bar{\nu}$	**JB test**
Inflation rate	3.192	2.099	1.224	4.274	82.498 (0.000)
Short-run expectations	3.446	2.190	1.542	5.234	157.160 (0.000)
Long-run expectations	3.052	1.508	1.203	3.327	63.840 (0.000)
Output gap	-0.248	2.254	-0.381	4.086	19.072 (0.000)

Note: JB test refers to the Jarque and Bera test for normality. Parentheses report p-values. All series reject normality at the 1% level.

The descriptive statistics of the inflation rate, the output gap, and inflation expectations over the analyzed period are reported in Table 3. As shown, the inflation rate has a positive mean $\bar{\mu}$ of 3.19%, reflecting the long-run average quarterly inflation rate in the sample. Its standard deviation $\bar{\sigma}$ of 2.10 indicates substantial unconditional volatility. Positive empirical skewness ($\bar{\zeta} = 1.22$) reveals asymmetry, while kurtosis ($\bar{\nu} = 4.27$) suggests a fat-tailed distribution consistent with occasional high-inflation episodes.

In contrast, the output gap has a mean close to zero ($\bar{\mu} = -0.25$), as expected given its definition as a cyclical deviation from potential output. Its empirical distribution is slightly negatively skewed ($\bar{\zeta} = -0.38$) and exhibits elevated kurtosis ($\bar{\nu} = 4.09$), indicating somewhat heavier tails than a Gaussian benchmark.

Turning to inflation expectations, short-run expectations display a higher mean ($\bar{\mu} = 3.45$) and greater dispersion ($\bar{\sigma} = 2.19$) than realized inflation. Their distribution is strongly right-skewed ($\bar{\zeta} = 1.54$) and leptokurtic ($\bar{\nu} = 5.23$), pointing to pronounced asymmetry and tail risk. By contrast, long-run expectations are more tightly anchored, with a lower standard deviation ($\bar{\sigma} = 1.51$) and kurtosis ($\bar{\nu} = 3.33$), and a mean ($\bar{\mu} = 3.05$) close to that of realized inflation. Their skewness ($\bar{\zeta} = 1.20$) remains positive but less pronounced than for short-run expectations.

The characteristics of the empirical distributions suggest that the observables are not normally distributed unconditionally. This is confirmed by the Jarque–Bera (JB) test, which rejects the null hypothesis of Gaussianity for each series.

3.3 Parameter estimates

This section reports the estimation results for the specifications summarized in Table 2 using short-run (S) and long-run (L) inflation expectations.

An important consideration in the use of survey-based inflation expectations concerns their measurement and informational content. As emphasized by Mavroeidis et al. (2014), survey expectations are potentially contaminated by a composite error term reflecting both measurement error and information frictions, arising from the reliance of survey respondents on a more limited information set than that assumed in the model, as well as their limited ability to filter out cost-push shocks. As a result, survey-based expectations are subject to noise and may not be strictly exogenous.

While lagging the expectations measure helps alleviate endogeneity concerns, it does not resolve the fundamental identification problem that arises when expectations exhibit limited variation. In such cases, their contribution to inflation dynamics becomes weakly identified, as emphasized by Mavroeidis et al. (2014) and discussed more broadly in Cochrane (2011). This implies that the empirical role of expectations may depend critically on the variability of the expectation proxy employed. In this context, alternative measures that are more strongly anchored, such as long-run expectations, may yield substantially different implications for both the estimated weight on expectations and the implied dynamics of the Phillips curve.

The estimated parameters and corresponding objective function values are reported in Tables 4 and 5. The aggregate loss and information criteria in Table 4 indicate that model fit improves substantially when moving from the baseline DMQ model to its PC-augmented counterparts, DMQ(S) and DMQ(L). The gain in model fit is more pronounced under the DMQ(S) specification, which is associated with steeper estimated slope parameter and larger expectation weights relative to the corresponding specification based on long-run expectations. This suggests that short-run inflation expectations provide stronger identifying variation in the estimated models, consistent with the identification issues discussed above. Whether the nearly flat slope parameters in the homogeneous DMQ(L) and DMQ-TV(L) specifications reflect weak identifying variation associated with long-run inflation expectations warrants further investigation.

Parameters governing the score-driven updating of the dynamic quantiles are broadly comparable across specifications, with some notable exceptions. First, the estimate of β under the benchmark DMQ model is close to unity ($\beta = 0.99999$), suggesting that the unconditional median contributes little information to the dynamic quantile updating process. In contrast, the lower estimates of β under the semi-structural specifications suggest that incorporating inflation expectations into the reference quantile process reduces the inertia, and hence the role of the backward-looking component. Second, when comparing specifications based on the two types of expectations, the higher estimates of β and α associated with long-run expec-

Table 4: Estimated parameters of the DMQ specifications with short-run (S) and long-run (L) inflation expectations over 1961Q1–2025Q4

	DMQ	**DMQ(S)**	**DMQ(L)**	**DMQ-TV(S)**	**DMQ-TV(L)**
			Dynamic quantile update		
α	0.531	0.459	0.526	0.471	0.572
β	1.000	0.883	0.941	0.925	0.940
γ	0.070	0.083	0.042	0.080	0.034
ϕ	0.998	0.997	0.998	0.997	0.998
			TVP-related parameters		
ρ_κ	–	–	–	0.013	0.281
ϕ_κ	–	–	–	0.756	0.998
ρ_β	–	–	–	0.088	0.302
ϕ_β	–	–	–	0.985	0.849
			Semi-structural parameters		
$1-\beta$	<0.001	0.117	0.059	0.075	0.060
κ	–	0.052	0.010	0.052	0.005
			Model fit		
Loss	6017.1	5886.2	5975.4	5871.3	5936.8
AIC	12042.2	11782.4	11960.9	11756.6	11887.6

Note: Under DMQ-TV, the reported β values correspond to sample averages. Loss values denote the final negative quasi-maximum likelihood (QML) objective function evaluated at the estimated parameter values.

tations indicate a more persistent dynamic quantile updating process, consistent with the lower variability and weaker identifying content of long-run inflation expectations. Third, the lower estimates of γ under the specifications with long-run expectations indicate a weaker response of the positive-valued processes around the reference quantile to new information, suggesting more muted distributional dynamics.

Distinct persistence dynamics also emerge in the TVP updating processes under the time-varying DMQ specifications. Specifically, the homogeneous slope dynamics become highly persistent, approaching a near random-walk process when long-run expectations are considered. In contrast, slope-updating persistence is less pronounced in the DMQ-TV(S) specification.

The opposite pattern emerges for the expectation-weight process, with stronger inertia in the relative weight assigned to short-run expectations compared with the specification based on long-run expectations.

A similar pattern arises in the dynamic quantile and TVP updating processes in the sDMQ specifications reported in Table 5. Notably, the average values of the slope parameters κ become more stable and more comparable across specifications. This suggests that removing the homogeneity constraint on the slope parameter alleviates potential identification issues associated with long-run inflation expectations. More importantly, the estimated slope parameters across the

Table 5: Estimated parameters of the sDMQ specifications with short-run (S) and long-run (L) inflation expectations over 1961Q1–2025Q4

	sDMQ(S)	**sDMQ(L)**	**sDMQ-TV(S)**	**sDMQ-TV(L)**
		Dynamic quantile update		
α	0.509	0.564	0.483	0.568
β	0.872	0.939	0.922	0.937
γ	0.084	0.025	0.075	0.024
ϕ	0.997	0.998	0.998	0.998
		TVP-related parameters		
ρ_κ	–	–	0.112	0.148
ϕ_κ	–	–	0.712	0.920
ρ_β	–	–	0.100	0.335
ϕ_β	–	–	0.985	0.849
ψ_s	–	–	0.969	0.839
		Semi-structural parameters		
$1-\beta$	0.128	0.117	0.078	0.063
κ	0.069	0.052	0.066	0.057
$\kappa_{\tau=0.05}$	0.092	0.093	0.091	0.081
$\kappa_{\tau=0.25}$	0.094	0.088	0.093	0.073
$\kappa_{\tau=0.50}$	0.077	0.081	0.076	0.067
$\kappa_{\tau=0.75}$	0.065	0.064	0.065	0.059
$\kappa_{\tau=0.95}$	0.003	0.003	0.005	0.006
		Model fit		
Loss	5845.6	5939.4	5831.1	5914.3
AIC	11709.1	11896.8	11676.1	11842.7

Note: Under the sDMQ specification, the reported values of κ are averages across quantiles, while under sDMQ-TV they are averaged across both time and quantiles. Under sDMQ-TV, both κ_p and β correspond to sample averages, and ϕ_β is obtained from DMQ-TV. The reported aggregated loss values correspond to the final negative quasi-maximum likelihood (QML) objective function evaluated at the estimated parameter values.

target probability levels reveal non-negligible heterogeneity, driven primarily by the substantially lower slope estimates in the upper tail of the conditional distribution.

Turning to the sDMQ-TV models, the high value of the shrinkage parameter ψ_s implies that the time- and quantile-averaged values of the time-varying slope parameter $\kappa_{p,t}$ remain comparable to those obtained under the static sDMQ specifications. Notably, the sDMQ-TV specification with short-run expectations attains the lowest aggregate loss and information criteria, indicating superior overall model fit.

Figure 2 provides a visual overview of the estimated slopes and reveals pronounced asymmetry. The slope is consistently steepest at the lower tail (long-run specifications) and the first quartile (short-run specifications) of the inflation distribution. This pattern suggests that inflation is more responsive to real activity under lower-tail and lower-quartile inflation outcomes. Moving toward the median, the slope flattens somewhat but remains moderately steep around the median, before declining sharply toward the upper tail, indicating a weak output-inflation trade-off at higher inflation outcomes.

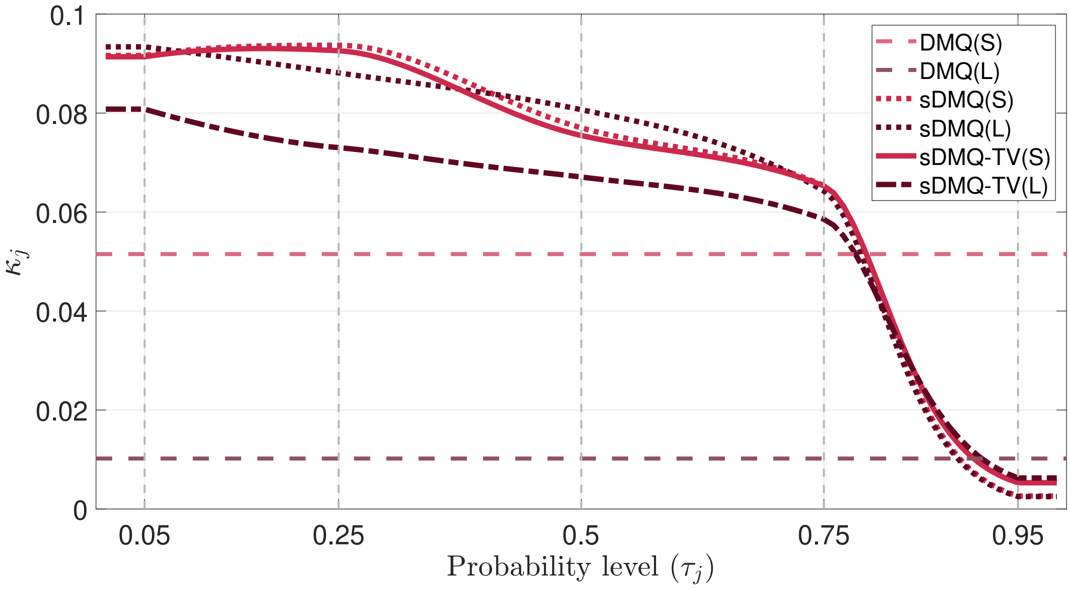

Figure 2: Phillips curve slopes across inflation quantiles obtained from different specifications. Values for the sDMQ-TV model correspond to sample-averaged slopes.

It should be noted that the static sDMQ model yields less stable estimates of the slope parameter in the lower tail, as reflected in their sensitivity to initial values and estimation settings. To address this, we conduct a robustness analysis in which the target tail quantiles are set to $\tau = 0.1$ and $\tau = 0.9$, with flat extrapolation applied beyond these thresholds to increase the effective number of observations in the tails. While the resulting estimates are qualitatively similar, we acknowledge the possibility of weak identification in the lower tail under the static sDMQ specification.

The ambiguity surrounding the static lower-tail slope parameter under the sDMQ specification likely reflects multiple factors, including the limited number of effective observations due to the rarity of deflationary episodes in the sample period and potential time variation. Given the improved model fit of the DMQ-TV and sDMQ-TV specifications relative to more restricted alternatives, the evidence points to time variation as an important contributing factor. Overall, these results indicate that the inflation–output gap trade-off differs systematically across the inflation distribution and evolves over time, highlighting the importance of modeling time-varying and quantile-heterogeneous slopes rather than relying on static homogeneous specifications.

In summary, the results indicate clear heterogeneity in the trade-off between inflation and the output gap across the inflation distribution. The static sDMQ model consistently reveals an asymmetric slope profile, with the steepest slopes at lower and central quantiles and a sharp decline toward the upper tail. In general, this pattern is preserved when allowing for time variation in the slope parameters. Model fit improves substantially under the sDMQ specifications relative to the benchmark DMQ models, with the sDMQ-TV specifications delivering the best overall fit.

3.4 Phillips curve slope dynamics

This section analyzes the evolution of the time-varying Phillips curve slope across quantiles of the conditional inflation distribution. By allowing the inflation–output gap trade-off to vary over time and across probability levels, we assess the extent of time variation and distributional asymmetry in the slope dynamics.

Figure 3 displays the time-varying Phillips curve slope estimates obtained from the sDMQ-TV specification with short-run inflation expectations, revealing important dynamics in the shape of the Phillips curve that remain hidden when considering its static distributional profile alone. The slope is steepest and exhibits the greatest variability at the first quartile and in the lower tail; it remains moderate and relatively stable around the median and declines toward higher probability levels, becoming nearly flat in the upper tail.

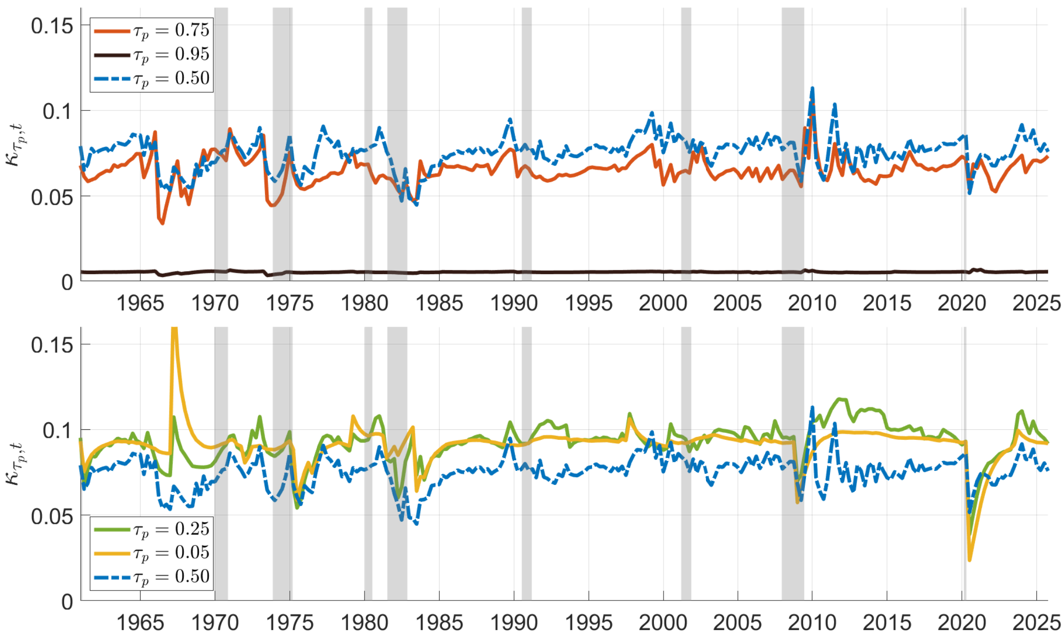

Figure 3: Time-varying Phillips curve slope parameters across the US conditional inflation distribution at the target probability levels obtained by sDMQ-TV specification with short-run expectations.

The results indicate an asymmetric Phillips curve across the conditional inflation distribution, with time-varying patterns that coincide with major macroeconomic episodes in the US economy. Importantly, this time variation alters the slope profile substantially in certain periods. For example, during the early 1980s, under Volcker monetary tightening, the slope asymmetry becomes more pronounced, with relatively steep slopes at lower and middle quantiles. As the monetary policy shock attenuates, however, the slope profile compresses, leading to more homogeneous slopes across quantiles, except for the upper tail.

Another notable change in the slope profile occurs in the post-GFC period, when the slope at the first quartile reaches historically high levels. This pronounced shift coincides with periods of effective lower bound and below-target inflation. From a New Keynesian perspective, this pattern is consistent with state-dependent Phillips curve dynamics, potentially reflecting greater inflation sensitivity to real marginal costs under low-inflation conditions and constrained monetary policy. Given the effective lower bound and the pronounced steepening at the first quartile, this pattern suggests that mechanisms beyond downward nominal rigidities were likely at work during this period, consistent with Fuhrer et al. (2012).

Lastly, following the COVID-19 pandemic, the Phillips curve relationship weakens markedly in the lower quantiles of the inflation distribution. This breakdown in the link between the output gap and inflation is consistent with the absence of disinflation. In contrast, standard New Keynesian models with a time-invariant Phillips curve slope would typically predict disinflation following a sharp but transitory contraction in output.

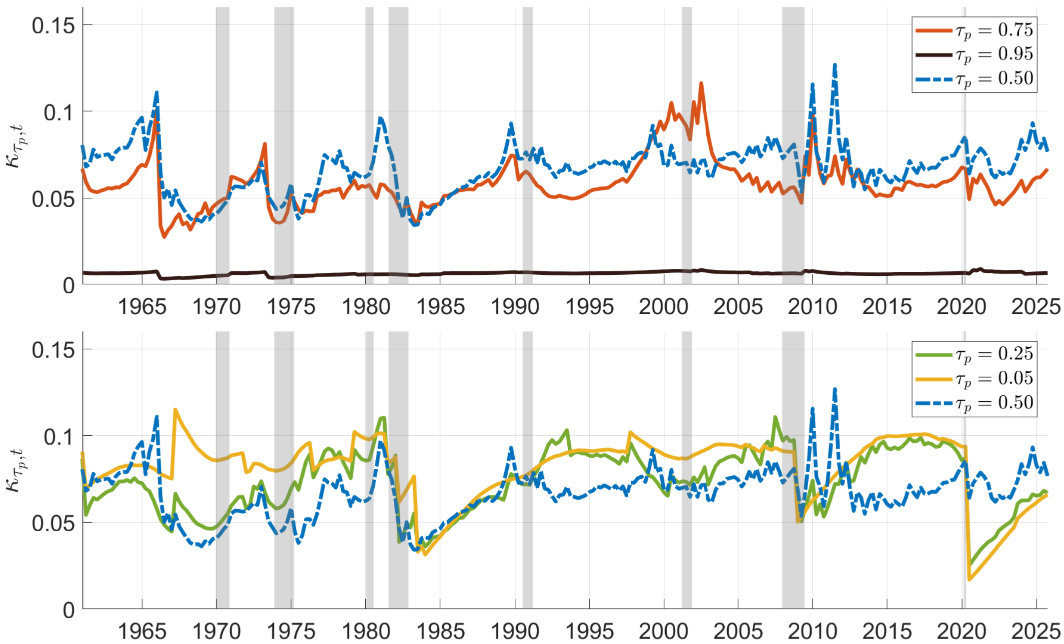

Figure 4: Time-varying Phillips curve slope parameters across the US conditional inflation distribution at the target probability levels obtained by sDMQ-TV specification with long-run expectations.

Figure 4 displays the estimated slopes based on the specification with long-run expectations. While the slope profile dynamics are broadly comparable to those observed under the short-run expectations specification, their time variation is more pronounced. Specifically, except for the post-Volcker disinflationary episode, the degree of heterogeneity exhibits greater variability and more substantial cross-quantile shifts over time. Importantly, the collapse of the Phillips curve relationship in the lower tail and at the first quartile during the post-COVID period becomes more evident relative to the specification with short-run expectations.

Overall, comparable Phillips curve dynamics emerge under both the long- and short-run expectations specifications, albeit with stronger persistence and greater variability under the long-run specification. The flattening of the Phillips curve in the lower tail and at the first quartile following the Great Recession and during the post-COVID period is consistent with the missing disinflation observed during these episodes. More broadly, the results highlight the importance of allowing for both time variation and distributional heterogeneity in the analysis of US inflation dynamics. The sDMQ-TV specification captures economically meaningful variation that is obscured in models imposing static or quantile-invariant slopes. The Phillips curve exhibits an asymmetric shape-profile across quantiles that evolves over time, with changes in both magnitude and shape across macroeconomic episodes. These findings suggest that conflicting empirical evidence on the Phillips curve flattening hypothesis may reflect substantial heterogeneity, asymmetry, and time variation across the inflation distribution.

3.5 Time-varying relative weight on inflation expectations

Figure 5 displays the time-varying weight on short-run and long-run inflation expectations, measured as $1 - \beta_t$, where β_t is the AR parameter governing inflation inertia. Higher values of $1 - \beta_t$ therefore indicate a stronger forward-looking component and a weaker role for backward-looking information in shaping inflation dynamics.

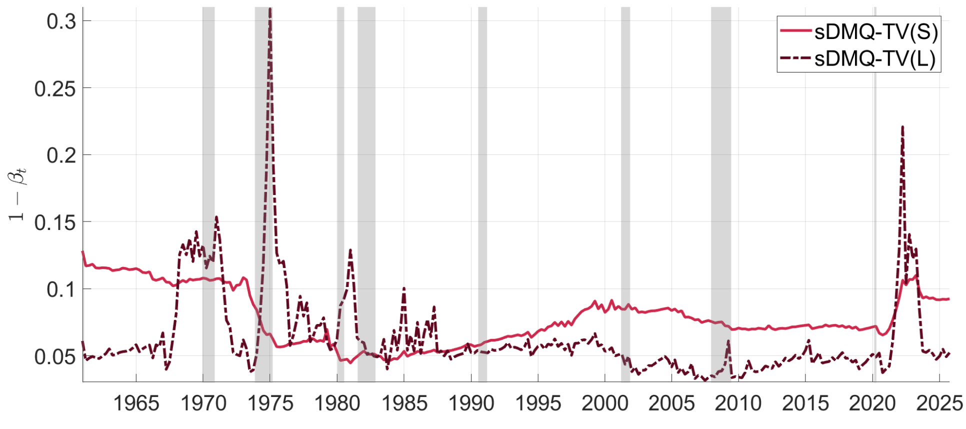

Figure 5: Time-varying relative weight on short-run and long-run forward-looking inflation expectation obtained from the sDMQ-TV specification.

The figure reveals regime dependence in the role of both short-run and long-run inflation expectations. In the late 1960s and early 1970s, the weight on expectations is relatively elevated, particularly for long-run expectations, consistent with weakly anchored expectations that amplified inflation dynamics and contributed to the Great Inflation. The fact that these periods are not systematically associated with steeper Phillips curve slopes supports the view that expectations played a central role in sustaining high inflation outcomes.

Comparing the two measures, the specification based on short-run expectations displays more persistent dynamics, while the specification based on long-run expectations exhibits sharper and more episodic movements, especially during periods of macroeconomic stress. Specifically, the weight on long-run expectations shows pronounced spikes during the 1970s inflationary episode and again in the post-2020 period, whereas the weight on short-run expectations evolves more gradually and remains at comparatively moderate levels. In the specification based on short-run expectations, the observed strong persistence with moderate time variation is consistent with earlier evidence showing that persistence in PCE and GDP deflator inflation remained comparatively high and stable since the 1960s (Pivetta and Reis, 2007; Benati, 2008; Wolters and Tillmann, 2015). This pattern suggests that price-setting behavior relies more heavily on long-run expectations during inflationary episodes, while the relative weight on short-run expectations remains more stable over time.

3.6 Dynamics of the US inflation quantiles

This section reports the dynamic conditional inflation quantiles for the sDMQ-TV models with both short-run and long-run inflation expectations.

By tracking the evolution of selected target quantiles, we analyze time variation in the conditional distribution of US inflation. Focusing on distributional dynamics rather than mean outcomes allows us to identify shifts in asymmetry, dispersion, and tail behavior across macroeconomic regimes.

The filtered conditional target inflation quantiles obtained with short-run expectations are reported in Figure 6. A clear pattern emerges in the shape of the distribution when comparing the five target quantiles over time. These distributional shifts are consistent with well-documented episodes in US macroeconomic history.

The oil shocks of the 1970s pushed the right tail of the inflation distribution markedly away from the median, consistent with elevated inflationary risk. The right tail asymmetry remained substantial for an extended period, reflecting strong inflation persistence and de-anchoring inflation expectations combined with a relatively weak sensitivity to output gap fluctuations above the central quantiles of the inflation distribution.

After the monetary tightening of the 1980s, consistent with consolidating monetary policy during the Great Moderation, the inflation quantiles became more tightly clustered, reflecting lower volatility during this interval. However, this period of relative stability was disrupted by the GFC. The widening spread between quantiles around the median reflects the increase in uncertainty during this period.

More recently, a clear contrast emerges between the pre- and post-GFC decades.

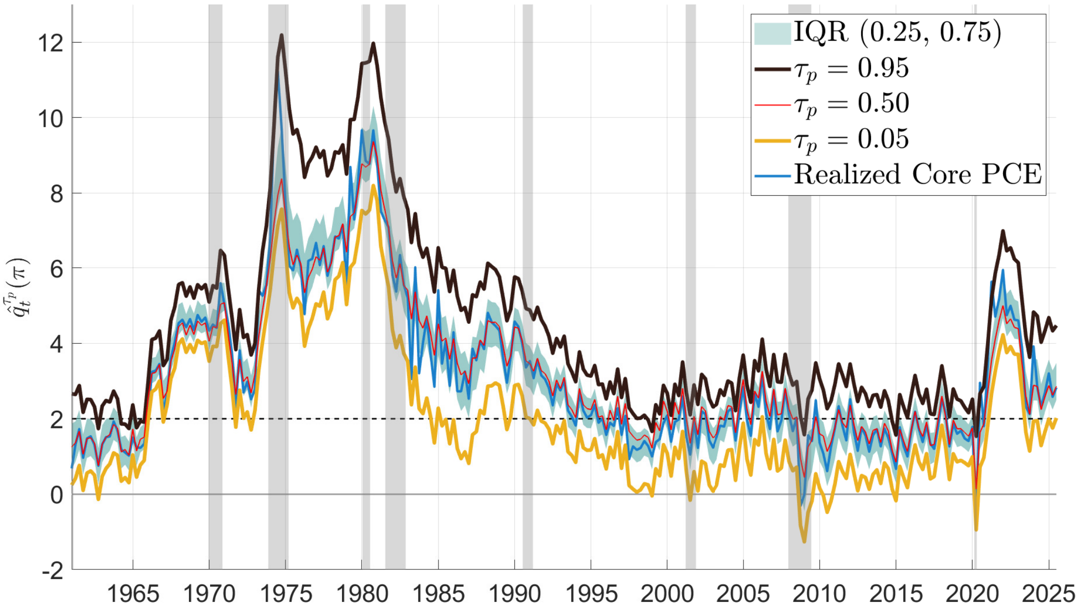

Figure 6: Filtered inflation quantiles obtained by the sDMQ-TV model with short-run expectations together with realized core CPI inflation. The horizontal dashed line represents the Fed's 2% PCE inflation target, shown for reference. Vertical shaded areas indicate recessions as identified by the National Bureau of Economic Research (NBER).

While realized PCE inflation after the GFC was only marginally lower than in the preceding period, the lower-tail quantiles at $\tau = 0.05$ periodically became negative. This indicates that, despite the Fed's ability to keep average inflation between 1% and its 2% PCE inflation target, deflationary risks remained non-negligible in the post-GFC environment, consistent with concerns raised by Fuhrer et al. (2012).

The severe macroeconomic shocks associated with the COVID-19 pandemic marked a turning point in inflation dynamics. After a brief period of downward pressure on prices, the subsequent demand surge ended a prolonged period characterized by relatively low inflationary risk. Importantly, the right-tail quantiles remained above 4%, indicating elevated inflation-at-risk at the end of the analyzed period. Moreover, the weak Phillips curve relationship in the upper tail, together with strong inflation persistence, helps explain the prolonged elevation of inflationary risk observed during this period.

The filtered inflation quantiles obtained under the long-run expectation specification reported in Figure 7 display broadly similar distributional dynamics, but with some notable differences relative to the short-run specification. Specifically, the long-run expectation specification generates more compressed dynamics during inflationary episodes. While the upper-tail inflation quantiles are generally less persistent, the around 4% inflation-at-risk at the end of the analyzed period remains robust when long-run expectations are considered.

Overall, the result demonstrates that the filtered conditional inflation quantiles exhibit strong persistence, and tend to diverge during periods of macroeconomic stress, especially in inflationary episodes. Importantly, the analysis of inflation

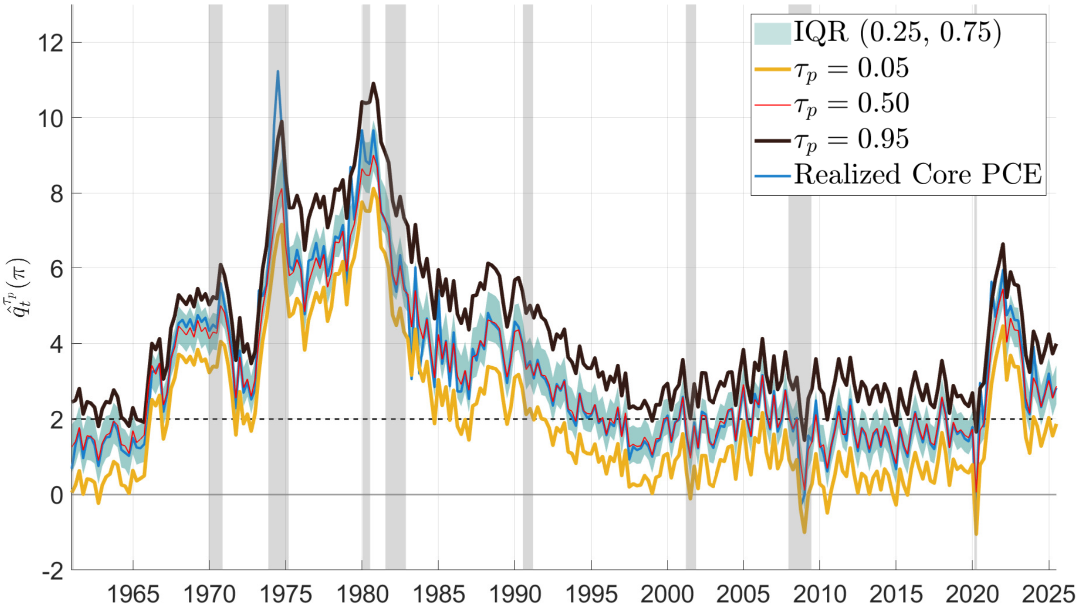

Figure 7: Filtered inflation quantiles obtained by the sDMQ-TV model with long-run expectations together with realized core CPI inflation. The horizontal dashed line represents the Fed's 2% PCE inflation target, shown for reference. Vertical shaded areas indicate recessions as identified by the National Bureau of Economic Research (NBER).

quantiles demonstrates that the sDMQ model effectively captures time-varying asymmetry and heteroskedasticity in the conditional distribution of the US inflation. The results show that inflation dynamics are asymmetric and evolve over time, with higher conditional variability in the right tail, particularly during recessions and high-inflation periods.

3.7 Time-varying conditional moments

The pronounced variability observed in the dynamic conditional inflation quantiles indicates that the US inflation distribution has undergone substantial changes over the past 65 years. However, direct quantitative interpretation of this variability based solely on the quantile dynamics is not straightforward.

To gain further insight, a skewed t-distribution, as proposed by Azzalini and Capitanio (2003), is fitted at each time t using the estimated target quantiles. The resulting implied conditional moments provide a convenient summary of the time-varying shape of the inflation distribution. Moreover, comparing the higher-order conditional moments obtained under specifications with short-run and long-run inflation expectations allows for a deeper assessment of the distributional implications of the differences in their dynamics.

The skew t density takes the form of

$$f(y;\mu,\sigma,\zeta,\nu) = \frac{2}{\sigma}\, d_T\left(\frac{y-\mu}{\sigma};\nu\right)\, T\left(\zeta\,\frac{y-\mu}{\sigma}\sqrt{\frac{\nu+1}{\nu+\left(\frac{y-\mu}{\sigma}\right)^2}}\,;\nu+1\right), \qquad (3.4)$$

where $d_T(\cdot)$ and $T(\cdot)$ denote the probability density function (PDF) and CDF of the skew t distribution, respectively. The parameters are location (μ), scale (σ), degrees of freedom (ν) governing tail thickness, and shape (ζ) governing the skewness. Estimating these parameters at each time t supports the assessment of the tail behavior of inflation by capturing its time-varying asymmetry and kurtosis implied by the estimated target quantiles.

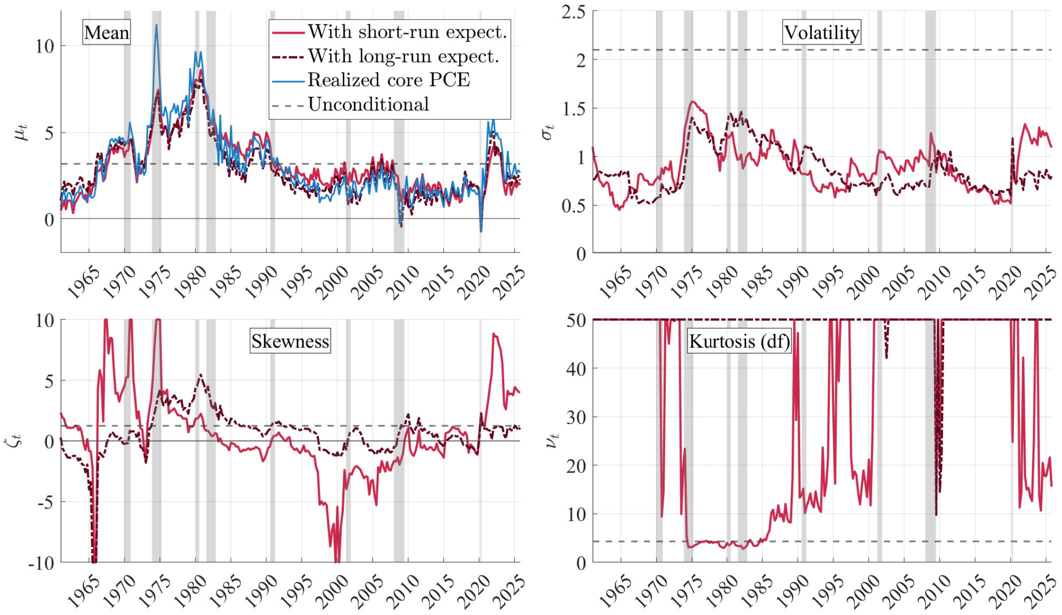

Figure 8: Filtered implied conditional moments obtained by fitting a skewed-t distribution, following Azzalini and Capitanio (2003), at each time t using the target quantiles. The parameters are the location (μ_t, upper left panel), the scale (σ_t, upper right panel), the shape parameter (ζ_t, lower left panel) governing skewness, and the degrees of freedom (ν_t, lower right panel) governing tail thickness. The horizontal dashed lines denote the corresponding unconditional moments. Shaded vertical areas indicate recessions as identified by the National Bureau of Economic Research (NBER).

Figure 8 displays the filtered time-varying conditional moments. The upper left panel shows the filtered conditional mean of inflation, which captures the low-frequency trend in core PCE by removing transitory shocks and extreme realizations. Realized core PCE exhibits pronounced short-run deviations around this mean, particularly during the 1970s inflation episode and the post-2020 inflation surge. The specification with short-run expectations captures more cyclical variation, whereas the long-run specification yields a smoother and more persistent path.

The upper right panel reports the filtered conditional inflation volatility, which peaks during the 1970s and early 1980s and declines markedly thereafter. Volatility remains elevated during the high-inflation period and persistently low during the Great Moderation, with only a modest increase following the GFC and in the post-2020 period. The difference between the two specifications becomes more pronounced: volatility is more time-varying and shock-sensitive under short-run expectations, but smoother and more persistent under long-run expectations due to anchoring.

A similar pattern emerges for the conditional skewness reported in the lower left panel. Under short-run expectations, skewness is strongly time-varying, with large positive spikes during the 1970s and occasional negative phases in the late 1960s and the late 1990s. In contrast, under long-run expectations the conditional inflation distribution remains close to symmetric from the 1980s onward. This pattern is consistent with the Great Moderation and reflects a transition associated with monetary policy consolidation, better-anchored inflation expectations, and reduced inflationary risks.

Lastly, the lower right panel reports filtered kurtosis across specifications. Under long-run expectations, excess kurtosis largely vanishes, whereas it remains pronounced under short-run expectations. This indicates that the specification with more anchored expectations yields a distribution closer to Gaussian, while short-run expectations generate heavier tails.

In sum, the results highlight substantial differences in the implied distributional dynamics across expectation proxies. The specification with short-run expectations yields richer time variation across all moments, with more pronounced volatility, skewness, and excess kurtosis. This stronger time variation reflects the transmission of greater short-run expectation variability into the conditional distribution. In contrast, the specification with long-run expectations produces a smoother and more stable distribution, characterized by attenuated higher-order moments and dynamics closer to a Gaussian benchmark. Overall, these findings underscore the important role of expectations in shaping the dynamics and higher-order properties of the conditional distribution of inflation.

4 Conclusions

This paper analyzed the US inflation dynamics by jointly modeling multiple quantiles within a hybrid New Keynesian Phillips curve framework. To this end, the Smoothed Dynamic Multiple-Quantile model was introduced, allowing for quantile-specific and time-varying parameters. By linking the time variation of a reduced set of quantiles to the full conditional distribution within a score-driven framework, the proposed specification captures heterogeneous responses of inflation to real activity across the distribution and over time.

The empirical results revealed pronounced time-varying and distributional nonlinearities in the inflation–output gap trade-off, indicating clear heterogeneity across the conditional inflation quantiles. Specifically, the Phillips curve exhibits an asymmetric profile across quantiles, with relatively steep slopes at lower probability levels and a sharp flattening towards the upper tail. Allowing for time variation showed that this distributional pattern evolves over time, consistent with state-dependent New Keynesian price setting behavior.

More broadly, the proposed framework provides a flexible and structurally interpretable tool for analyzing inflation dynamics across different parts of the distribution and over time. The results highlight the central role of inflation expectations in shaping high-inflation episodes, particularly during the Great Inflation, while emphasizing the dominance of persistence and weak real activity pass-through in the lower quantiles following the Global Financial Crisis and during the COVID-19 pandemic. The combination of a flattening Phillips curve slope in the lower quantiles and strong persistence helps explain the absence of disinflation during sharp output contractions, while persistently flat slopes in the upper tail explain the challenge of stabilizing inflation when inflationary risks are elevated.

In addition, the results underscore that the choice of expectation proxy plays a critical role for identification, with short-run expectations revealing richer distributional dynamics and more pronounced higher-order moments, while more anchored long-run expectations yield smoother distributions but stronger time variation in the cross-quantile Phillips curve shape.

Lastly, the filtered inflation quantiles provide a quantitative measure of inflationary risk, capturing both its persistence and asymmetric evolution over time. Notably, the upper tail conditional inflation quantile has remained around 4% since the COVID-19 pandemic indicating that inflationary risk remains a non-negligible factor at the end of the analyzed period. Importantly, the weak Phillips curve relationship in the upper tail, together with strong inflation persistence, helps explain the prolonged elevation of inflationary risk observed during this period.

Overall, the results suggest that exploiting distributional information and time variation can provide additional insights into inflation dynamics beyond those available from conventional limited-information empirical approaches.

More generally, the framework is not limited to inflation, and its generality allows it to be applied in other macro-financial contexts where time-varying quantile-heterogeneous parameters, persistence, and tail risks are central. In addition, its ability to model time-varying higher-order moments enables full distributional forecasting, providing a more comprehensive assessment of tail risks.

References

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

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

Ball, L. and S. Mazumder (2011). Inflation Dynamics and the Great Recession. *Brookings Papers on Economic Activity 42*(1 Spring), 337–405.

Benati, L. (2008). Investigating inflation persistence across monetary regimes. Technical Report Working Paper No. 851, European Central Bank.

Benigno, P. and G. B. Eggertsson (2023). It's baaack: The surge in inflation in the 2020s and the return of the non-linear phillips curve. Working Paper 31197, National Bureau of Economic Research.

Blasques, F., C. Francq, et al. (2023). Quasi score-driven models. *Journal of Econometrics 231*(2), 235–257.

Catania, L. and A. Luati (2023). Semiparametric modeling of multiple quantiles. *Journal of Econometrics 237*(2, Part B), 105365.

Clark, T. E. and T. Doh (2014). Evaluating alternative models of trend inflation. *International Journal of Forecasting 30*(3), 426–448.

Cochrane, J. H. (2011). Determinacy and identification with taylor rules. *Journal of Political Economy 119*(3), 565–615.

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795.

Fritsch, F. N. and R. E. Carlson (1980). Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis 17*(2), 238–246.

Fuhrer, J. C., G. P. Olivei, and G. M. B. Tootell (2012). Inflation dynamics when inflation is near zero. *Journal of Money, Credit and Banking 44*(s1), 83–122.

Furceri, D., D. Giannone, M. F. Kisat, M. W. R. Lam, and H. Li (2025, May). Debt-at-risk. IMF Working Papers 2025/086, International Monetary Fund.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Society Monographs. Cambridge University Press.

Hazell, J., J. Herreño, E. Nakamura, and J. Steinsson (2022). The slope of the Phillips curve: Evidence from u.s. states. *The Quarterly Journal of Economics 137*(3), 1299–1344.

Koenker, R. and F. Machado, José A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association 94*(448), 1296–1310.

Komunjer, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics 128*(1), 137–164.

Kotz, S., J. Kozubowski, Tomasz and K. Podgórski (2001). *The Laplace Distribution and Generalizations*. Boston: Birkhäuser.

López-Salido, D. and F. Loria (2024). Inflation at risk. *Journal of Monetary Economics 145*, 103570.

Mavroeidis, S., M. Plagborg-Møller, and J. H. Stock (2014, March). Empirical evidence on inflation expectations in the new keynesian phillips curve. *Journal of Economic Literature 52*(1), 124–88.

Negro, M. D., D. Giannone, M. P. Giannoni, and A. Tambalotti (2017). Safety, liquidity, and the natural rate of interest. *Brookings Papers on Economic Activity*, 235–294.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861–1957. *Economica 25*(100), 283–299.

Pivetta, F. and R. Reis (2007). The persistence of inflation in the United States. *Journal of Economic Dynamics & Control 31*(4), 1326–1358.

Poiraud-Casanova, S. and C. Thomas-Agnan (2000). About monotone regression quantiles. *Statistics & Probability Letters 48*(1), 101–104.

Rossi, B., A. Inoue, and Y. Wang (2024, Jun). Has the phillips curve flattened? French Stata Users' Group Meetings 2024 22, Stata Users Group.

White, H., T.-H. Kim, and S. Manganelli (2015). Var for var: Measuring tail dependence using multivariate regression quantiles. *Journal of Econometrics 187*(1), 169–188.

Wolters, M. H. and P. Tillmann (2015). The changing dynamics of US inflation persistence: a quantile regression approach. *Studies in Nonlinear Dynamics & Econometrics 19*(2), 161–182.