Kothe, Rafael

Regime-dependent predictive accuracy and structural stability of Eurozone inflation swaps

Primary publication

Kothe, Rafael (2026): Regime-dependent predictive accuracy and structural stability of Eurozone inflation swaps, in: Economics letters, Amsterdam [u.a.]: Elsevier, Vol. 262, No. 112826, pp. 1–4, doi: 10.1016/j.econlet.2026.112826.

Economics Letters 262 (2026) 112826

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Regime-dependent predictive accuracy and structural stability of Eurozone inflation swaps

Rafael Kothe *

Institute of Economics, University of Bamberg, Feldkirchenstr. 21, 96052, Bamberg, Germany
Bamberg Research Training Group on Bounded Rationality, Heterogeneity, and Network Effects, University of Bamberg, Feldkirchenstr. 21, 96052, Bamberg, Germany
Bamberg Graduate School of Social Sciences, University of Bamberg, Feldkirchenstr. 21, 96052, Bamberg, Germany

ARTICLE INFO

JEL classification:
C22
C52
E52
E31
G12

Keywords:
Inflation expectations
Inflation-linked swaps
Markov-switching models
Regime dependence
Forecast accuracy
Inflation risk premia

ABSTRACT

This paper investigates the predictive accuracy of Eurozone inflation-linked swaps (ILS) across volatility regimes using a Markov-switching framework and regime-specific Mincer–Zarnowitz regressions. Results show a sharp divergence by maturity. While 12-month ILS remain approximately unbiased, their forecast precision (RMSE) deteriorates sharply in high-volatility states. In contrast, longer maturities (24–36 months) develop statistically significant, large positive biases (up to 378 basis points) and calibration losses evident across both volatility periods — a finding masked by standard asymptotic inference. These findings highlight the structural, persistent nature of the mispricing at medium horizons and the risk of policy misinterpretation.

1. Introduction

The anchoring and predictive power of inflation expectations are central to effective monetary policy, especially in the Eurozone. They shape wage- and price-setting and, along with other macroeconomic indicators, influence the European Central Bank's (ECB) monetary policy conduct and ability to achieve its target (Baumann et al., 2021). Inflation-linked swaps (ILS) are a primary, high-frequency market-based measure of these expectations (Burban et al., 2024).

The extant literature provides mixed, horizon-dependent evidence. While short-term swaps can be nearly unbiased under benign conditions (Campbell et al., 2023), their predictive content at longer horizons and their performance relative to professional surveys remain debated (Acharya et al., 2024). Recent macroeconomic shocks have further exposed weaknesses in ILS performance under volatile conditions (Cecchetti et al., 2022). The absence of a systematic, regime-dependent analysis of ILS degradation presents a significant research gap, hindering policy interpretation.

I apply a Markov-Switching Autoregressive (MS-AR(2)) model (Hamilton, 1989) to distinguish high- and low-volatility regimes (2008–2024) and use Mincer–Zarnowitz regressions (Mincer and Zarnowitz, 1969) to examine forecast unbiasedness. Predictive power is highly conditional on both the volatility regime and contract maturity. While 12-month swaps remain approximately unbiased, their forecast accuracy deteriorates sharply in high-volatility states, with Root Mean Squared Errors (RMSE) rising by nearly 278%. In contrast, contracts with longer maturities (24–36 months) develop economically and statistically significant biases that are structurally evident across both regimes, with intercept deviations in turbulent periods reaching up to 378 basis points (bps) and RMSE rising by 92%–195% relative to tranquil periods. These results structurally reconcile previously conflicting evidence and have direct implications for monetary policy interpretation of market-based expectations.

2. Data, structural identification and conditional testing

2.1. Data and variable construction

I use monthly Eurozone zero-coupon inflation swap (ZCIS) rates at the 12-, 24-, and 36-month maturities from the London Stock Exchange Group. The 12-month series spans from July 2008 to January 2024,

* Correspondence to: Institute of Economics, University of Bamberg, Feldkirchenstr. 21, 96052, Bamberg, Germany.
E-mail address: rafael.kothe@uni-bamberg.de.

https://doi.org/10.1016/j.econlet.2026.112826
Received 12 November 2025; Received in revised form 12 January 2026; Accepted 18 January 2026
Available online 4 February 2026

while the 24- and 36-month series end in January 2023 and January 2022, respectively. The predictor is the n-month ZCIS rate at time t, $r_t(n)$. Realized inflation, $\pi_t(n)$, is computed as the ex post annualized growth of the Harmonised Index of Consumer Prices excluding Tobacco (HICPxT), applying the standard three-month indexation lag. The HICPxT series is the revised Eurostat dataset obtained via FRED. This series is not a real-time vintage; results therefore reflect ex post predictive accuracy, necessitating caution when translating the specific numerical magnitudes of bias to real-time policy evaluation. Exact formulas, construction details, and robustness to real-time vintages are reported in the Online Appendices B

2.2. First stage: Volatility regime identification

Inflation volatility regimes are identified using the MS-AR(2) model applied to the year-on-year growth of HICPxT. The model distinguishes a persistent low-volatility regime and a shorter high-volatility regime, which differ primarily in variance. This regime classification serves as an exogenous input for the subsequent Mincer–Zarnowitz forecast evaluation. The formal model specification, parameter estimates, and robustness diagnostics are reported in Appendix B, with summary results in Table B1.

2.3. Second stage: Conditional forecast evaluation

To assess how predictive performance varies, I analyze Mincer–Zarnowitz (MZ) regressions conditional on the prevailing volatility regime, $s_t \in \{\text{low}, \text{high}\}$. The following specification is estimated separately for observations within each state:

$$\pi_t(n) = \alpha_{s_t} + \beta_{s_t} r_t(n) + \epsilon_{t,s_t}. \tag{1}$$

where unbiasedness and full information content correspond to the joint null hypothesis $H_0 : (\alpha_{s_t}, \beta_{s_t}) = (0, 1)$. Standard errors for Eq. (1) are estimated using a Newey–West HAC estimator. Given the known unreliability of asymptotic inference in small, persistent high-volatility samples ($N \approx$ 33–55), I formally validate all key hypothesis tests using a Moving Block Bootstrap (MBB) to generate robust small-sample p-values. All reported p-values are subsequently adjusted for multiple comparisons across maturities and regimes using the Benjamini–Hochberg procedure. The methodology and detailed comparative analysis are provided in the Online Appendices C.4.1 and C.4.3.

3. Empirical results and discussion

3.1. Accuracy and bias by horizon and state

The primary analysis relies on MZ regressions (Eq. (1)) to formally test for bias and calibration. Throughout this section, I define the forecast error ($FE_{t,n}$) as the difference between realized inflation and the swap rate: $FE_{t,n} = \pi_t(n) - r_t(n)$. Figs. 1(a) and 1(b) illustrate the estimated coefficients (α and β) and their standard 95% HAC-based confidence intervals. Although these figures visualize the economic magnitudes of the regime-switch, the confidence intervals are based on asymptotic HAC errors, which (as noted in Section 2.3 and the Online Appendix C.4) are unreliable in small, persistent samples.[1]

Therefore, the primary conclusions on statistical significance rely on the robust Moving Block Bootstrap (MBB) p-values, corrected for multiple hypothesis testing (see Table C8). The analysis confirms that performance degrades sharply with both maturity and volatility.

Table 1
Cross-regime bias and predictive accuracy of inflation swaps.

Maturity	Bias (α) [bps]		RMSE(fe) [pp]	
	Low Vol.	High Vol.	Low Vol.	High Vol.
12-Month	22.3	151.6	0.793	2.997
24-Month	119.9	350.1	1.132	3.336
36-Month	249.3	377.9	1.523	2.917
Max HV bias (≥24 m)	378 bps			
RMSE(fe) increase (≥24 m)	91%–195%			

Notes: Bias is computed as the Mincer–Zarnowitz intercept (α) multiplied by 100 to express values in basis points (bps). RMSE(fe) is the root mean squared error of the raw forecast error (realized inflation minus swap rate) in percentage points (pp). "Max HV bias" is the largest absolute high-volatility bias among maturities ≥24 months. "RMSE(fe) increase" reports the minimum and maximum percentage increase from low- to high-volatility regimes for the same maturities.

For short-horizon (12 m) swaps, the breakdown is identified as a loss of precision, not a formal bias. The point estimates for the intercept remain small (Fig. 1(a)). Crucially, after applying robust multiple-testing corrections, the 12-month swap passes all formal tests for bias and efficiency in the high-volatility regime. The individual intercept test is not significant (BH-adjusted $p = 0.544$), and the joint hypothesis of full unbiasedness ($H_0 : (\alpha, \beta) = (0, 1)$) also cannot be rejected (BH-adjusted $p = 0.544$). This confirms that the 12-month swap's breakdown is primarily a loss of precision, not a statistically significant bias or calibration failure, dominated by a sharp drop in precision: the Root Mean Squared Error (RMSE) rises by approximately 278%.

The lack of a statistically significant intercept contrasts with the significant mean bias identified by the HAC t-tests. This is not a contradiction but a consequence of the regression's lower statistical power in the high-volatility subsample. High inflation volatility inflates the standard errors of the regression coefficients, making a rejection of the null hypothesis ($\alpha = 0$) statistically less likely, even when the economic magnitude of the bias is large.

Third, contracts with longer maturities (24–36 months) exhibit a structural breakdown. The 24-month tenor highlights the critical importance of the multiple-testing correction. While the high-volatility regime shows a definitive bias (BH-adjusted $p < 0.001$), the low-volatility regime does not. The raw bootstrap test for bias in the low-vol regime ($p = 0.038$) is corrected to non-significance by the BH procedure (BH-adjusted $p = 0.057$), filtering out what would have been a false positive. The test for the slope coefficient ($H_0 : \beta = 1$) is also informative: it fails to reject in the high-vol regime (BH-adjusted $p = 0.594$) but rejects in the low-vol regime (BH-adjusted $p = 0.042$), indicating a significant calibration failure even in tranquil periods.

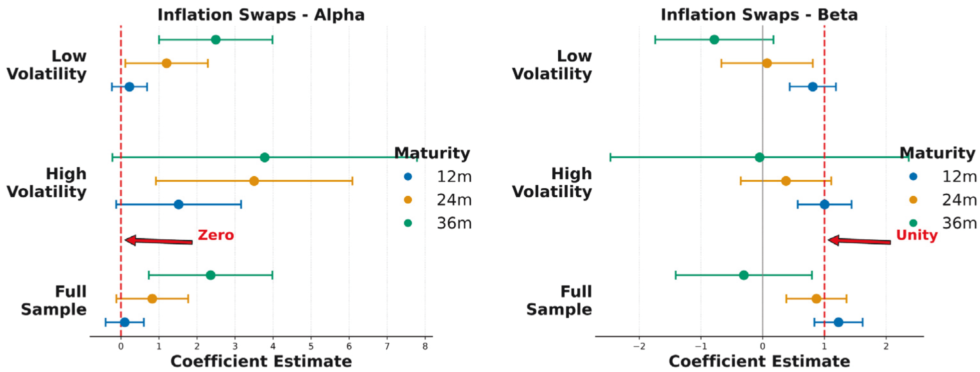
The 36-month tenor, in contrast, suffers a robust structural failure. For the high-volatility regime, the estimated bias (intercept) of 378 basis points (Fig. 1(a) and Table 1) is statistically definitive, a finding masked by standard inference (HAC $p = 0.064$) but confirmed by the robust, multiple-testing corrected inference (BH-adjusted $p < 0.001$). Furthermore, the slope coefficient collapse, illustrated in Fig. 1(b), is also statistically robust in both the high- and low-volatility regimes, with the test strongly rejecting $H_0 : \beta = 1$ (BH-adjusted $p < 0.001$ in both). This confirms a total breakdown of the contract's predictive properties, regardless of the volatility state. Table 1 summarizes the economic magnitudes of these effects.

The complete Mincer–Zarnowitz evaluation, from standard HAC results (Appendix C) to the robust bootstrap methodology (Appendix C.4) and final multiple-testing corrections (Appendix C.4.3), is fully documented in the Online Appendix.

3.2. Mechanisms: IRP and technical frictions

These regime-dependent patterns suggest a joint role for time-varying inflation risk premia (IRP) and technical frictions (Pflueger and Viceira, 2011). The full empirical analysis, detailed in Online

[1] A preliminary analysis of mean forecast errors (see Online Appendix C.1, Figure C2 and Table C3) confirms that while mean errors are significant in high-volatility, simple HAC-robust t-tests (e.g., 12-month: $p = 0.022$) are misleading. As the main text shows, these findings are a statistical artifact of the test's low power, which is corrected by the more robust joint Mincer–Zarnowitz bootstrap.



(a) Intercept coefficients (α). (b) Slope coefficients (β).

Fig. 1. Estimated Mincer–Zarnowitz Coefficients by Regime and Maturity. Panel (a) displays the intercepts (α), which test for systematic bias. Panel (b) displays the slopes (β), which test for calibration and predictive power. Estimates are shown for the full sample and conditional on the low- and high-volatility regimes. Whiskers represent 95% asymptotic HAC-based confidence intervals, which are shown for illustrative purposes but are unreliable in small samples (see Section 2.3). Markers are color-coded by maturity: 12-month (blue), 24-month (orange), and 36-month (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Appendix C.5, confirms this and reveals a sharp, maturity-dependent distinction. The 36-month tenor's failure is strongly associated with the IRP mechanism, consistent with a highly significant interaction between the high-volatility regime and risk aversion (proxied by the CBOE Volatility Index, VIX; MBB $p < 0.001$). In contrast, the 12-month breakdown is a different phenomenon, characterized by a structural level-shift (HighVol dummy, MBB $p \leq 0.017$) and a direct pass-through of energy shocks that is robustly identified only when controlling for credit risk (MBB $p = 0.024$).

The findings provide critical context to the market-versus-survey debate. Acharya et al. (2024), for example, find that 1-year ILS recently outperformed survey forecasts in a simple RMSE comparison, precisely because swaps responded faster to the post-pandemic inflation shock. My results are consistent with this: I also find that 1-year ILS remain robustly unbiased (all 12 m BH-adj. $p > 0.39$). The "faster response" they identify corresponds to the sharp 278% increase in RMSE I document, which I characterize as a loss of precision, not a loss of unbiasedness. This finding, however, creates a potential policy trap if this 1-year reliability is incorrectly generalized across the term structure.

4. Conclusion: Policy dualism and structural caveats

The predictive performance of Eurozone inflation swaps is highly conditional on maturity and volatility regime. Short-horizon ILS lose precision but remain approximately unbiased during volatile conditions, passing formal joint tests (where $H_0 : (\alpha, \beta) = (0, 1)$ is not rejected). In contrast, my central finding is that this reliability does not extend to longer tenors. For the 36-month swap, I find a complete structural failure, with statistically definitive biases (BH-adj. $p < 0.001$) and a collapse in calibration (BH-adj. $p < 0.001$) in both volatility regimes. This failure is structural, not a high-volatility artifact.

This state-dependent reliability points toward a dual policy approach: monitoring short-term swaps directly, while adjusting for the systematic bias in swaps with longer maturities (24–36 months), which prove unreliable even in calm regimes. The primary risk is misinterpreting this bias – driven by IRP and market dysfunction – as de-anchored expectations, precisely because these instruments are treated as key longer-term market-based indicators for inflation expectations. This argues for regime-sensitive evaluation frameworks that account for such nonlinearities.

Code availability

The code used to generate the results and perform the analyses in this study is not publicly available but is available from the corresponding author upon reasonable request.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author utilized OpenAI's models, specifically GPT-4 and GPT-5, to enhance the readability and language of the manuscript. All content was subsequently reviewed and edited by the author, who assumes full responsibility for the final version of the publication.

Funding

The author of this work is supported by a doctoral scholarship from the Hans Böckler Foundation, Germany and their financial support is gratefully acknowledged. The views expressed in this research are solely those of the author and do not necessarily reflect the opinions of the Hans Böckler Foundation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

An earlier version of this paper was presented at the 29th Annual Workshop on Economics with Heterogeneous Interacting Agents (London, 2025), the Bavarian Macro Day (Nuremberg, 2025), the International Workshop on Economic Complexity and Macroeconomic Dynamics (Valencia, 2024), the 10th Meeting of the German Network for New Economic Dynamics (Gießen, 2024), and the 27th Annual Workshop on Economic Science with Heterogeneous Interacting Agents (Bamberg, 2024).

I thank Johannes Pfeifer and Lisa Sheenan for invaluable comments on the econometric methodology; Filippo Gusella and Sylvain Barde for discussions on Markov-switching models; and Lebogang Mateane and Anthony M. Diercks for their feedback on earlier drafts.

I appreciate the comments of all workshop participants, particularly Thomas Lux, Corrado Di Guilmi, Tomasz Makarewicz, Philipp Mundt, and Armin Aminian. Finally, I am especially grateful to my academic supervisor, Frank Westerhoff, for his continuous guidance.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.econlet.2026.112826.

Data availability

The datasets analyzed during the current study are proprietary to LSEG and cannot be made publicly available due to licensing restrictions. Access to these data can be obtained via LSEG subject to their terms and conditions.

References

Acharya, S., Harikrishnan, N., Mertens, T.M., et al., 2024. Evaluating Forecast Performance of Market-Based Measures of Inflation Expectations in Europe. FEDS Notes Board of Governors of the Federal Reserve System, Washington, D.C., http://dx.doi.org/10.17016/2380-7172.3550.

Baumann, U., Darracq Paries, M., Westermann, T., et al., 2021. Inflation Expectations and their Role in Eurosystem Forecasting. Occasional Paper Series No. 264, European Central Bank, Frankfurt am Main.

Burban, V., De Backer, B., Vladu, A.L., 2024. Inflation (de-) Anchoring in the Euro Area. Working Paper Series, No 2964, European Central Bank, Frankfurt am Main.

Campbell, C., Diercks, A.M., Sharpe, S.A., et al., 2023. The Swaps Strike Back: Evaluating Expectations of One-Year Inflation Board of Governors of the Federal Reserve System (U.S.). Finance and Economics Discussion Series No. 2023-061, Washington, D.C., http://dx.doi.org/10.17016/FEDS.2023.061.

Cecchetti, S., Grasso, A., Pericoli, M., 2022. An Analysis of Objective Inflation Expectations and Inflation Risk Premia. Temi di discussione (Economic working papers) 1380, Bank of Italy, Economic Research and International Relations Area, Rome.

Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57 (2), 357–384. http://dx.doi.org/10.2307/1912559.

Mincer, J.A., Zarnowitz, V., 1969. The evaluation of economic forecasts. In: Mincer, J.A. (Ed.), Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance. National Bureau of Economic Research (NBER), New York, pp. 3–46.

Pflueger, C.E., Viceira, L.M., 2011. Inflation-indexed bonds and the expectations hypothesis. Annu. Rev. Financ. Econ. 3 (1), 139–158. http://dx.doi.org/10.1146/annurev-financial-102710-144843.