SIMPLE TESTS FOR THE CORRECT SPECIFICATION OF CONDITIONAL PREDICTIVE DENSITIES

2025

BANCO DE ESPAÑA
Eurosistema

Documentos de Trabajo
N.º 2535

Gergely Ganics and Lluc Puig Codina

SIMPLE TESTS FOR THE CORRECT SPECIFICATION OF CONDITIONAL PREDICTIVE DENSITIES

SIMPLE TESTS FOR THE CORRECT SPECIFICATION OF CONDITIONAL PREDICTIVE DENSITIES (*)

Gergely Ganics (**)

BANCO DE ESPAÑA

Lluc Puig Codina (***)

UNIVERSITY OF ALICANTE AND BANCO DE ESPAÑA

(*) We thank the Editor Luis Julián Álvarez, an anonymous referee, Juan Mora López, Gabriel Pérez Quirós, Barbara Rossi, Tatevik Sekhposyan, and seminar participants at the SNDE 2025, CESC 2025, Banco de España and IAAE 2025 for very useful comments and suggestions. The views expressed herein are those of the authors, and do not necessarily reflect those of the Banco de España or the Eurosystem.

(**) gergely.ganics@bde.es.

(***) lluc.puig@ua.es, lluc.puig@bde.es.

Documentos de Trabajo. N.º 2535

September 2025

https://doi.org/10.53479/40825

ISSN: 1579-8666 (online edition)

Abstract

We propose a simplified framework for evaluating conditional predictive densities based on the probability integral transform (PIT). The approach accommodates a wide range of estimation schemes, including expanding and rolling windows, and applies to both stationary and non-stationary processes. By treating the PIT as a primitive, our approach enables researchers to apply widely used tests in settings where their validity was previously uncertain. Monte Carlo simulations demonstrate favorable size and power properties of the tests. In an empirical application, we show that incorporating stochastic volatility into an unobserved components model is essential for generating correctly calibrated density forecasts of US industrial production growth at both monthly and quarterly frequencies.

Keywords: predictive density, forecast evaluation, probability integral transform, Kolmogorov–Smirnov test, Cramér–von Mises test.

JEL classification: C22, C52, C53.

Resumen

Proponemos un marco simplificado para evaluar densidades predictivas condicionadas basado en la transformación integral de probabilidad (TIP). Este planteamiento acomoda un amplio abanico de métodos de estimación, incluyendo ventanas móviles y expansivas, y es adaptable tanto a procesos estacionarios como no estacionarios. Al tratar la TIP como un elemento primitivo, nuestra estrategia permite a los investigadores realizar contrastes de especificación comúnmente utilizados a contextos donde su validez era, previamente, desconocida. Las simulaciones Montecarlo muestran propiedades favorables en cuanto al tamaño y potencia de las pruebas. En una aplicación empírica mostramos que incorporar volatilidad estocástica en un modelo de componentes no observados es esencial para generar densidades predictivas correctamente especificadas de la producción industrial de Estados Unidos, tanto a frecuencias mensuales como trimestrales.

Palabras clave: densidad predictiva, evaluación de predicciones, transformación integral de probabilidad, test de Kolmogorov–Smirnov, test de Cramér–von Mises.

Códigos JEL: C22, C52, C53.

1 Introduction

Since the turn of the millennium, probabilistic forecasting in economics and finance has grown from a niche area into a central component of forecasting practice across academia, policy institutions, and industry. See Timmermann (2000) for early applications, and Gneiting and Katzfuss (2014) and Timmermann and Elliott (2016) for more recent overviews. Researchers and policymakers now increasingly focus not only on the expected path of variables like inflation, growth, or house prices, but also on the uncertainty around these forecasts—see, for example, the Federal Open Market Committee's Summary of Economic Projections[1] or the International Monetary Fund's house prices-at-risk analysis (International Monetary Fund, 2019).

Given the importance of probabilistic forecasts, assessing their correct specification is essential. A realization may fall in a low-density region of the forecast, which can result from misspecification or the realization of a low-probability event. Statistical inference helps distinguish between these possibilities. For this purpose, Diebold et al. (1998) introduced the probability integral transform (PIT) into economics and proposed several visual assessments of two key properties of the PIT under correct specification: uniformity and independence. To handle parameter estimation uncertainty, Bai (2003) applied Khmaladze's transformation to eliminate nuisance parameters that appear in the empirical process once estimation uncertainty is accounted for, obtaining a Kolmogorov-type test. Parameters are evaluated at their pseudo-true limiting values. Corradi and Swanson (2006a,c) developed and extended tests based on the PIT empirical process by incorporating dynamic misspecification and parameter estimation error.

Multivariate extensions using the Rosenblatt transformation can be found in Diebold et al. (1999), Bai and Chen (2008) and Dovern and Manner (2020).

Knüppel (2015) formulated tests of PIT uniformity based on raw moments. More recently, Knüppel et al. (2023) developed tests of correct calibration of multivariate predictive distributions based on proper scoring rules. Both aforementioned papers treat forecasts as primitives, without accounting for parameter estimation uncertainty, akin to Diebold et al. (1998).

Rossi and Sekhposyan (2019) proposed tests which preserve parameter estimation error asymptotically under the null hypothesis of correct calibration. Users can readily apply their method in a variety of settings where correcting for parameter estimation error would be either undesirable, such as when one is interested in actual forecasting performance (as opposed to population-level performance), or infeasible. Typical cases of the latter are survey forecasts, and

[1]The complete history of the Federal Open Market Committee's Summary of Economic Projections is available at https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm.

in general when the researcher evaluating the predictions does not have access to the model which generated them. Their test is also one of the few ones applicable to multiple-step-ahead predictive densities, as most of the econometrics literature on specification testing of conditional predictive densities has focused on one-step-ahead forecasts. Prominent applications of Rossi and Sekhposyan (2019) include Adrian et al. (2019), Adrian et al. (2021), Lhuissier (2022) and Clark et al. (2024). All these applications share a common element: the models were estimated in expanding windows, which is not covered by Rossi and Sekhposyan (2019), whose tests were designed for a finite (*i.e.*, rolling) estimation window. In particular, their Assumption 2 allows the predictive distribution to depend only on a finite lag of observations. This stands in contrast to the usual practice of estimating models with dynamic latent variables—such as dynamic factor models, hidden Markov models, or stochastic volatility models—where filtering typically exploits the entire history of the process. Thus, bridging the gap between the theoretical validity of Rossi and Sekhposyan's (2019) tests and their empirical applications is of immediate interest.

Our first contribution is providing an alternative set of assumptions under which we prove that the tests remain valid. By abstracting from parameter estimation uncertainty and treating the PITs as primitives (similarly to Diebold et al., 1998, Knüppel, 2015, Knüppel et al., 2023 and Demetrescu et al., 2025), our framework accommodates both rolling and expanding window estimation schemes, as well as stationary and non-stationary processes. As a novelty, we consider weighted test statistics, which allow researchers to focus on specific parts of the predictive distribution (*e.g.*, the left tail) without completely disregarding incorrect calibration appearing in other regions. Second, we demonstrate through Monte Carlo simulations that the tests have favorable size and power properties under our proposed alternative assumptions. Finally, in an empirical application forecasting US industrial production, we find that incorporating stochastic volatility into an unobserved components model leads to correctly calibrated density forecasts.

Two remarks are in order. First, in line with Rossi and Sekhposyan (2019), our framework is useful to evaluate *forecasts*, not *models* (see also Diebold, 2015). Second, in contrast to *e.g.*, Amisano and Giacomini (2007) and Gneiting and Ranjan (2011) but similarly to earlier contributions cited above which build on the PIT, we focus on *absolute*, rather than *relative*, density forecast evaluation.

The remainder of the paper is structured as follows. Section 2 introduces the notation and defines the main objects. Section 3 presents our proposed tests. Section 4 demonstrates the size and power of our testing procedure through Monte Carlo simulations, and Section 5 illustrates our approach through an empirical example of forecasting US industrial production. Section 6 concludes. Proofs and additional results are collected in the Appendices.

2 Notation and definitions

This section introduces the notation and definitions describing the econometric framework. The notation largely follows that of Rossi and Sekhposyan (2019).

Consider the stochastic process $\{\boldsymbol{Z}_t : \Omega \to \mathbb{R}^{n+1}\}_{t=1}^{\infty}$ defined on a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Partition the observed vector $\boldsymbol{Z}_t$ as $(y_t, \boldsymbol{X}_t')'$ where $y_t : \Omega \to \mathbb{R}$ is the scalar variable of interest and $\boldsymbol{X}_t : \Omega \to \mathbb{R}^n$ is a set of predictors, potentially incorporating lagged values of y_t. Let $\mathscr{F}_t = \sigma(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_t)$ denote the natural filtration.

Our interest lies in the true but unknown h-step-ahead predictive density of y_{t+h} conditional on $\mathscr{F}_t$, denoted by $\phi_0(\cdot \mid \mathscr{F}_t)$, and the corresponding conditional predictive cumulative distribution function (CDF), denoted by $\Phi_0(\cdot \mid \mathscr{F}_t)$.[2] We assume that the forecaster has at her disposal a sample of size $T+h$ from the above stochastic process, divided into an initial in-sample period of size E, fixed and finite, and an out-of-sample period of size P, where $E + P - 1 + h = T + h$. The forecaster obtains a sequence of the h-step-ahead ($h < \infty$ and fixed) predictive CDFs conditional on the σ-field $\mathscr{I}_t \subseteq \mathscr{F}_t$ (one can think of $\mathscr{I}_t$ as the forecaster's information set), denoted by $\Phi_{t+h}(y \mid \mathscr{I}_t)$.[3] No restrictions are imposed on the methods used by the forecaster, be it classical, Bayesian, some heuristic or a combination of these (*e.g.*, when the researcher applies judgment to adjust a model-based forecast). Researchers are free to use both rolling and expanding window estimation schemes, making this test a pseudo out-of-sample test.

The probability integral transform (PIT) is the conditional predictive CDF evaluated at the realization, formally:

$$\xi_{t+h} \equiv \Phi_{t+h}(y_{t+h} \mid \mathscr{I}_t) = \int_{-\infty}^{y_{t+h}} \phi_{t+h}(y \mid \mathscr{I}_t)\, \mathrm{d}y\,. \tag{1}$$

For $r \in [0,1]$, let $\varphi(r) \equiv \mathbb{P}(\xi_{t+h} \leq r) - r$ be the difference between the CDFs of the PIT and the uniform distribution. Let $\mathbb{1}\{\cdot\}$ denote the indicator function and consider $\varphi_P(r)$ given by

$$\varphi_P(r) \equiv \frac{1}{P} \sum_{t=E}^{T} \left(\mathbb{1}\{\xi_{t+h} \leq r\} - r\right)\,. \tag{2}$$

Finally, let $\Psi_P(r)$ be the corresponding uniform empirical process:

$$\Psi_P(r) \equiv \sqrt{P}\varphi_P(r) = P^{-1/2} \sum_{t=E}^{T} \left(\mathbb{1}\{\xi_{t+h} \leq r\} - r\right)\,. \tag{3}$$

[2]The true conditional density may depend on the forecast horizon h (which is taken as fixed) and time period t, which we omit without loss of generality to simplify notation. Furthermore, $\phi_0(\cdot \mid \cdot)$ and $\Phi_0(\cdot \mid \cdot)$ denote generic densities and distributions, respectively, not necessarily Gaussian.

[3]As before, $\Phi_{t+h}(\cdot \mid \cdot)$ stands for a generic distribution, not necessarily Gaussian.

3 Asymptotic tests

This section presents the proposed tests of correct calibration of predictive densities and the underlying assumptions. Often, prominent papers using the PIT to test the specification of predictive distributions, such as Corradi and Swanson (2006a) and Rossi and Sekhposyan (2019), specify a series of assumptions about the behavior of the stochastic process $\{\boldsymbol{Z}_t\}$, such as α-mixing at a particular rate, and prove that some of these desirable characteristics are inherited by the PITs. This paper takes a different approach, akin to Diebold et al. (1998), Knüppel (2015) and Knüppel et al. (2023), in that we directly study the behavior of the PITs. By leveraging the results in Lemmas 1 and 2 below, we obtain the same asymptotic distribution of $\Psi_P(r)$ as Rossi and Sekhposyan (2019) in an alternative framework. On the one hand, our approach abstracts from parameter estimation uncertainty and the specific estimation scheme (*e.g.*, rolling or recursive window) that a researcher might choose. However, as long as a forecaster is comfortable with the assumptions listed in this section, our results provide theoretical support for the continued use of Rossi and Sekhposyan's (2019) test even in cases which were not covered by their original contribution. For convenience, all proofs are collected in Appendix A.

The null hypothesis states that all predictive distributions are well specified, as in Rossi and Sekhposyan (2019), formally:

$$H_0 : \Phi_{t+h}(y \mid \mathscr{I}_t) = \Phi_0(y \mid \mathscr{F}_t) \;\text{ for all }\; t = E, \ldots, T\,. \tag{4}$$

The alternative hypothesis, H_A, is the negation of H_0. The null hypothesis can be interpreted as testing that the forecaster could not have specified more accurate predictive densities exploiting all available information.

Our proposed framework relies on the following two assumptions.[4]

Assumption 1 (Continuity). $\Phi_0(y_{t+h} \mid \mathscr{F}_t)$ *is continuous.*

Assumption 2 (Pairwise stationarity). $\mathbb{P}(\Phi_0(y_{t+h} \mid \mathscr{F}_t) \leq r_1, \Phi_0(y_{t+h+d} \mid \mathscr{F}_{t+d}) \leq r_2) = F_d(r_1, r_2)$, *where* $r_1, r_2 \in [0, 1]$.

These assumptions are similar to Rossi and Sekhposyan's (2019) Assumption 1 (*ii*) and (*iii*). Assumption 1 rules out CDFs of discrete (or certain mixed) distributions. Assumption 2 imposes a pairwise stationarity condition on the PITs. Importantly, the joint distribution of the PITs d

[4] With a slight abuse of notation, the subindices of y refer to the time indices of the CDFs, with y being the argument, not the realization.

periods apart is only allowed to depend on d but not on t. Note that in the terminology of Sklar (1959), $F_d(\cdot,\cdot)$ is the bivariate copula of $(y_{t+h} \mid \mathscr{F}_t)$ and $(y_{t+h+d} \mid \mathscr{F}_{t+d})$. As an example, take the case of a Gaussian random walk. The h-step-ahead forecast errors constitute a moving average of order $h-1$ with all coefficients being 1. As a result, these are jointly normal with correlation $\rho_{h,d} = \left|\frac{h-d}{h}\right|$ if $|d| \leq h-1$, and 0 otherwise. It means that, for a fixed h, $F_d(r_1, r_2)$ is a Gaussian copula with correlation parameter $\rho_{h,d}$. The process satisfies Assumption 2 despite being integrated of order one, and thus being outside the coverage of Assumption 1 (i) of Rossi and Sekhposyan (2019). The basis of the test is the uniformity of the PITs, summarized in Lemma 1.

Lemma 1 (Conditional uniformity of PITs)**.** *Under H_0 and Assumption 1, the PITs are uniformly distributed:* $\{\xi_{t+h}\}_{t=E}^{T} \sim U(0,1)$.

In the case of $h = 1$, it is known that the PITs are independently and identically distributed (*iid*), see Lemma 1 of Bai (2003), so the classical Central Limit Theorem (CLT) can be applied. In the general case, PITs are $(h-1)$-dependent, as shown in Lemma 2, to which a corresponding CLT applies.

Lemma 2 (PITs are $(h-1)$-dependent)**.** *Under H_0 and Assumption 1, the PITs $\{\xi_{t+h}\}_{t=E}^{T}$ are $(h-1)$-dependent.*

The asymptotic distribution of the uniform empirical process, as the size of the out-of-sample period diverges ($P \to \infty$), is given by Theorem 1.

Theorem 1 (Asymptotic distribution of $\Psi_P(r)$)**.** *Under H_0 and Assumptions 1 and 2,*

$$\Psi_P(\cdot) \Rightarrow \Psi(\cdot)\,, \tag{5}$$

where "$\Rightarrow$" denotes weak convergence, $\Psi(\cdot)$ is a mean-zero Gaussian process with covariance kernel $\sigma(r_1, r_2) = \sum_{d=1-h}^{h-1} (F_d(r_1, r_2) - r_1 r_2)$, and $\mathbb{P}(\Psi(\cdot) \in C[0,1]) = 1$, where $C[0,1]$ is the space of continuous functions on the unit interval.

The covariance kernel will typically contain nuisance parameters. In the special case of $h = 1$ no nuisance parameters appear as the variables in F_0 are identical, and $\sigma(r_1, r_2) \equiv \min\{r_1, r_2\} - r_1 r_2$. Hence, $\Psi(\cdot)$ corresponds to a Brownian bridge process. Note that our covariance kernel is written as a truncated version of the one in Rossi and Sekhposyan's (2019) Theorem 3. However, by Lemma 2, which also applies to their case, $F_d(r_1, r_2) = r_1 r_2$ for $|d| > h-1$, hence the covariance kernels are in fact identical.

To obtain the critical values of several empirical CDF-based test statistics (described shortly) in the presence of nuisance parameters ($h > 1$), we propose a slightly modified weighted block bootstrap procedure following that of Rossi and Sekhposyan (2019).[5] Given the sample ω, let $\xi_{t+h}(\omega)$ denote a particular realization of the PIT. Consider the following *bootstrap* empirical process:

$$\Psi_P^*(r;\omega) = P^{-1/2} \sum_{t=E}^{T-\ell+1} \eta_t \sum_{i=t}^{t+\ell-1} \left(\mathbb{1}\left\{ \xi_{i+h}(\omega) \leq r \right\} - r \right) . \tag{6}$$

The relevant assumption on the simulated auxiliary random variable η_t is the following.

Assumption 3 (Bootstrap). $\{\eta_t\}_{t=E}^{T-\ell+1}$ *are independent random variables, also independent of* $\{\xi_{t+h}\}_{t=E}^{T}$, *have zero mean, variance* $1/\ell$ *and* $\mathbb{E}\left(\eta_t^4\right) = \mathcal{O}(1/\ell^2)$, *where* $\ell \to \infty$ *as* $P \to \infty$, *and* $\ell = o\big(P^{1/2}\big)$.

The validity of the bootstrap is given in the following theorem.

Theorem 2 (Bootstrap validity). *Under* H_0 *and Assumptions 1 to 3,*

$$\Psi_P^*(\cdot;\omega) \Rightarrow \Psi(\cdot), \;\; \omega - a.s. \tag{7}$$

As suggested by Inoue (2001), an appealing choice of the auxiliary variables is $\eta_t \overset{iid}{\sim} \mathcal{N}(0, 1/\ell)$, as the finite-sample distribution of $\Psi_P^*(\cdot;\omega)$ will be Gaussian.

Due to $\ell \to \infty$ as $P \to \infty$ in Assumption 3, the block size will asymptotically be larger than $h - 1$, the window where the PITs exhibit dependence as per Lemma 2, and so it will capture all relevant dependencies. However, in practical applications, depending on the choice of ℓ, this might not be the case. When the estimated covariance kernel is missing additional terms, it will induce size distortion. The bootstrap will tend to deliver lower (higher) critical values if the excluded bivariate copulas $F_d(\cdot,\cdot)$ exhibit positive (negative) dependence. Jointly with the condition $\ell = o\big(P^{1/2}\big)$, the previous remark implies that applications where h is large will require a much larger P. In Section 4 we provide some examples where this behavior arises.

By Theorems 1 and 2 and the Continuous Mapping Theorem (CMT), one obtains the asymptotic behavior of several empirical CDF-based test statistics, summarized in Corollary 1. We consider both unweighted and weighted Kolmogorov–Smirnov- and Cramér–von Mises-type test statistics, denoted by κ_P and C_P, respectively, although the researcher is free to use any other to which the CMT applies. In particular, the weighted test statistics emphasize regions of interest through the non-negative continuous weight function $w(r)$, similarly to the methods proposed

[5] In turn adapted from Inoue (2001).

by Corradi and Swanson (2006b) and Gneiting and Ranjan (2011) for relative density forecast evaluation. In this paper, we consider the following weight functions (taken from Gneiting and Ranjan, 2011), which put emphasis on the regions in parentheses: *i*) $w(r) = (1-r)^2$ (left tail), *ii*) $w(r) = r^2$ (right tail), *iii*) $w(r) = r(1-r)$ (center), *iv*) $w(r) = (2r-1)^2$ (tails). The unweighted statistics considered by Rossi and Sekhposyan (2019) obtain by setting $w(r) = 1$. Corollary 1 also covers the case if a researcher is interested in testing correct specification in a certain region only (*e.g.*, in the left tail, $r \in [0, 0.1]$), rather than over the full distribution.[6]

Corollary 1 (Asymptotic distribution of Kolmogorov–Smirnov- and Cramér–von Mises-type test statistics)**.** *Under H_0 and Assumptions 1 to 3:*

$$\kappa_P \equiv \sup_{r\in[0,1]} |\Psi^*(r;\omega)w(r)| \xrightarrow{d} \sup_{r\in[0,1]} |\Psi(r)w(r)| \ \ \omega - a.s. \tag{8}$$

$$C_P \equiv \int_0^1 \Psi_P^*(r;\omega)^2 w(r)\,\mathrm{d}r \xrightarrow{d} \int_0^1 \Psi(r)^2 w(r)\,\mathrm{d}r \ \ \omega - a.s. \tag{9}$$

The bootstrap simulation algorithm is described in Appendix B.

The results in this section show that the tests of Rossi and Sekhposyan (2019) are valid under an alternative set of assumptions. Most importantly, at the cost of treating the PIT as a primitive and abstracting from parameter estimation error, one no longer needs to rely on the α-mixing assumption on the data $\{\mathbf{Z}_t\}$, that the conditional predictive distribution only depends on a fixed set of R past and present observables, and that the model is estimated using a rolling window of size R. Therefore, we provide the necessary theoretical guarantees for a wide range of applications of Rossi and Sekhposyan's (2019) tests, whose validity had previously been uncertain.

For one-step-ahead forecasts, the test statistics' distribution is free of nuisance parameters, hence it is possible to tabulate their critical values. In Panels A and B, respectively, Table 1 shows the critical values for Kolmogorov–Smirnov-type and Cramér–von Mises-type tests in various regions and with various weight functions (in rows) at significance levels of 1%, 5% and 10%, for various sample sizes, denoted by P, and asymptotically (in columns). Researchers can use these to test the correct specification of one-step-ahead predictive densities, similarly to Rossi and Sekhposyan's (2019) Table 1, but our table *i*) also includes critical values for various finite sample sizes, *ii*) covers weighted tests as well, and *iii*) uses slightly different quantiles to define tails (0.1 and 0.9 versus their 0.25 and 0.75, respectively).

[6]This region, denoted by ρ, is allowed to be a finite union of closed sets on $[0, 1]$, not including empty sets or singletons. In this case one takes the supremum over ρ or integrates over ρ in Equations (8) and (9), respectively.

Table 1: Critical values for tests of correct specification of one-step-ahead predictive densities

Panel A: Kolmogorov–Smirnov-type tests

Region or weight function		$P=25$ 1%	5%	10%	$P=50$ 1%	5%	10%	$P=100$ 1%	5%	10%	$P=200$ 1%	5%	10%	Asymptotic 1%	5%	10%
Full	$r \in [0,1]$	1.58	1.32	1.19	1.59	1.33	1.20	1.60	1.34	1.20	1.61	1.34	1.20	1.61	1.34	1.21
Left tail	$r \in [0,0.1]$	0.90	0.63	0.52	0.86	0.65	0.57	0.84	0.66	0.58	0.85	0.66	0.58	0.83	0.66	0.58
Right tail	$r \in [0.9,1]$	0.90	0.63	0.52	0.86	0.65	0.57	0.84	0.66	0.58	0.85	0.66	0.58	0.83	0.66	0.58
Left half	$r \in [0,0.5]$	1.50	1.24	1.10	1.52	1.24	1.11	1.53	1.25	1.11	1.53	1.26	1.12	1.53	1.25	1.12
Right half	$r \in [0.5,1]$	1.50	1.24	1.10	1.52	1.24	1.11	1.53	1.25	1.11	1.53	1.26	1.12	1.53	1.25	1.12
Center	$r \in [0.1,0.9]$	1.58	1.32	1.19	1.59	1.33	1.20	1.60	1.34	1.20	1.61	1.34	1.20	1.61	1.34	1.21
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.97	0.74	0.62	0.94	0.73	0.64	0.92	0.73	0.65	0.92	0.74	0.65	0.90	0.74	0.66
Left tail	$w(r)=(1-r)^2$	0.85	0.66	0.58	0.83	0.67	0.59	0.82	0.67	0.60	0.82	0.67	0.60	0.81	0.67	0.60
Right tail	$w(r)=r^2$	0.85	0.66	0.58	0.83	0.67	0.59	0.82	0.67	0.60	0.82	0.67	0.60	0.81	0.67	0.60
Center	$w(r)=r(1-r)$	0.37	0.31	0.27	0.38	0.31	0.28	0.38	0.31	0.28	0.38	0.31	0.28	0.38	0.31	0.28
Tails	$w(r)=(2r-1)^2$	0.73	0.57	0.50	0.70	0.56	0.49	0.67	0.55	0.49	0.66	0.54	0.49	0.64	0.54	0.49

Panel B: Cramér–von Mises-type tests

Region or weight function		$P=25$ 1%	5%	10%	$P=50$ 1%	5%	10%	$P=100$ 1%	5%	10%	$P=200$ 1%	5%	10%	Asymptotic 1%	5%	10%
Full	$r \in [0,1]$	0.73	0.46	0.35	0.74	0.46	0.35	0.74	0.46	0.35	0.74	0.46	0.35	0.74	0.46	0.35
Left tail	$r \in [0,0.1]$	0.31	0.14	0.09	0.28	0.15	0.11	0.26	0.15	0.11	0.26	0.15	0.11	0.26	0.15	0.11
Right tail	$r \in [0.9,1]$	0.31	0.14	0.09	0.28	0.15	0.11	0.26	0.15	0.11	0.26	0.15	0.11	0.26	0.15	0.11
Left half	$r \in [0,0.5]$	0.84	0.51	0.38	0.85	0.51	0.38	0.85	0.51	0.38	0.85	0.51	0.38	0.85	0.51	0.38
Right half	$r \in [0.5,1]$	0.84	0.51	0.38	0.85	0.51	0.38	0.85	0.51	0.38	0.85	0.51	0.38	0.85	0.51	0.38
Center	$r \in [0.1,0.9]$	0.89	0.56	0.42	0.90	0.56	0.42	0.91	0.56	0.42	0.90	0.56	0.42	0.91	0.56	0.42
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.22	0.12	0.09	0.20	0.12	0.09	0.19	0.12	0.09	0.19	0.12	0.09	0.18	0.12	0.10
Left tail	$w(r)=(1-r)^2$	0.23	0.14	0.11	0.23	0.14	0.11	0.23	0.14	0.11	0.23	0.14	0.11	0.23	0.14	0.11
Right tail	$w(r)=r^2$	0.23	0.14	0.11	0.23	0.14	0.11	0.23	0.14	0.11	0.23	0.14	0.11	0.23	0.14	0.11
Center	$w(r)=r(1-r)$	0.16	0.10	0.07	0.16	0.10	0.07	0.16	0.10	0.07	0.16	0.10	0.07	0.16	0.10	0.07
Tails	$w(r)=(2r-1)^2$	0.13	0.08	0.06	0.13	0.08	0.06	0.13	0.08	0.06	0.13	0.08	0.06	0.13	0.08	0.06

Note: The table reports the critical values of the correct specification of predictive densities in various regions and with various weights (in rows) at the 1%, 5% and 10% significance levels for different sample sizes, denoted by P, and asymptotically (in columns), based on Kolmogorov–Smirnov-type and Cramér–von Mises-type statistics in Panels A and B, respectively. The finite-sample critical values are simulated based on the independent standard uniform distribution of the PITs. The asymptotic critical values are simulated based on the limiting Gaussian process. The number of Monte Carlo simulations is 1,000,000. The domain of r is discretized over $[0:0.001:1]$.

4 Monte Carlo results

In this section, we analyze the size and power properties of our test in finite samples, using a variety of DGPs (Data Generating Processes). We repeat the Monte Carlo exercises in Rossi and Sekhposyan (2019) with the difference that the true predictive conditional CDF, denoted by $\Phi_0(y_{t+h}|\mathscr{F}_t)$, is allowed to be a function of all past values instead of only the last R values.[7] In their notation, we allow for $R/P \to 1$. Additionally, we extend the Monte Carlo experiments to include non-linear and non-Gaussian processes.[8] In the exercises, the predictive distribution used by the forecaster coincides with the true predictive distribution under H_0, despite parameters being estimated. Therefore we still consider a simple rather than a composite hypothesis. In all exercises, the nominal size of the tests is 5%.

[7]We adapt DGPs S1–S5 and P1–P4 from Rossi and Sekhposyan (2019).

[8]DGPs S6–S7 and P5–P6 are novel relative to Rossi and Sekhposyan (2019).

4.1 Size analysis

First, we investigate the size properties of the tests. DGPs S1–S3 are based on one-step-ahead forecast densities, while DGPs S4–S7 correspond to multi-step-ahead forecasts.

DGP S1: Baseline Model. The first model features a time-varying intercept μ_t and a non-persistent predictor x_{t-1}, and is given by

$$y_t = \mu_t + x_{t-1} + \varepsilon_t \,, \tag{10}$$

$$\mu_t = \left(\frac{\sum_{j=2}^{t-1} x_{j-1} y_j}{\sum_{j=2}^{t-1} x_{j-1}^2} - 1 \right) x_{t-1} \,, \tag{11}$$

where $x_t \overset{iid}{\sim} \mathcal{N}(0,1)$, $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$, and x_t and ε_t are independent contemporaneously and at all leads and lags. The forecasting model is $y_t = \beta x_{t-1} + e_t$ with $e_t \overset{iid}{\sim} \mathcal{N}(0,1)$, where β is recursively estimated by OLS.

DGP S2: Extended Model. The second model adds persistence to the regressor, and reads

$$y_t = \mu_t + \gamma x_{t-1} + \varepsilon_t \,, \tag{12}$$

$$x_t = 0.2 + 0.8 x_{t-1} + \nu_t \,, \tag{13}$$

$$\mu_t = \left(\frac{\sum_{j=2}^{t-1} x_{j-1} y_j}{\sum_{j=2}^{t-1} x_{j-1}^2} - \gamma \right) x_{t-1} \,, \tag{14}$$

where $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$, $\nu_t \overset{iid}{\sim} \mathcal{N}(0,1.08^2)$, ε_t and ν_t are independent contemporaneously and at all leads and lags, and $\gamma = 0.48$. The forecasting model is $y_t = \beta x_{t-1} + e_t$ with $e_t \overset{iid}{\sim} \mathcal{N}(0,1)$, where β is recursively estimated by OLS. The parametrization, taken from Rossi and Sekhposyan (2019), mimics a scenario where a researcher forecasts one-quarter-ahead US real GDP growth with the lagged term spread from 1959:Q1 to 2010:Q3.

DGP S3: GARCH(1,0) Model. Introduced by Engle (1982) and generalized by Bollerslev (1986), Generalized Autoregressive Conditional Heteroskedasticity or GARCH models have become a cornerstone of financial econometrics. The third model is a GARCH$(1,0)$, written as

$$y_t = \sigma_t \varepsilon_t \,, \tag{15}$$

$$\sigma_t^2 = \rho_{0,t} + \rho_1 \sigma_{t-1}^2 \,, \tag{16}$$

$$\rho_{0,t} = \left(\frac{1}{t-2} \sum_{j=2}^{t-1} (\rho_{0,j} + \rho_1 \sigma_{j-1}^2) \varepsilon_j^2 - \rho_1 \sigma_{t-1}^2 \right) \,, \tag{17}$$

where $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$ and $\rho_1 = 0.1$. The forecasting model is $y_t = \sqrt{\gamma_t} e_t$ where $\gamma_t = (t-2)^{-1} \sum_{j=2}^{t-1} y_j^2$ and $e_t \overset{iid}{\sim} \mathcal{N}(0,1)$.

DGP S4: Serial Correlation Model. Adding a moving average (MA) component to the error term in DGP S1, the data are generated by

$$y_t = \mu_t + x_{t-1} + \varepsilon_t + \rho\varepsilon_{t-1}\,, \tag{18}$$

where $x_t \overset{iid}{\sim} \mathcal{N}(0,1)$, $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$, x_t and ε_t are independent contemporaneously and at all leads and lags, $\rho = 0.2$, and μ_t is defined as in DGP S1. The forecasting model is $y_t = \beta x_{t-1} + e_t$ with $e_t \overset{iid}{\sim} \mathcal{N}(0, 1+\rho^2)$, where β is recursively estimated by OLS.

DGP S5: IMA Model. The integrated moving average (IMA) DGP is given by

$$y_t = \mu_t + \varepsilon_t - \sum_{j=1}^{h-1} \rho^j \varepsilon_{t-j}\,, \tag{19}$$

$$\mu_t = \frac{1}{t-2} \sum_{j=2}^{t-1} y_j\,, \tag{20}$$

where $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, 1.261^2)$, $\rho = 0.275$, and h corresponds to the forecast horizon. This parametrization was chosen by Rossi and Sekhposyan (2019) from Table 3 in Stock and Watson (2007), who modeled changes in GDP price inflation during the Great Inflation, between 1960:Q1 and 1983:Q4. The estimated model is $y_t = \beta + e_t$ with $e_t \overset{iid}{\sim} \mathcal{N}(0, (1 + \sum_{j=1}^{h-1} \rho^{2j})1.261^2)$, where β is recursively estimated by OLS.

Even though the models for multi-step-ahead forecasts, DGPs S4 and S5, are written in terms of one-step-ahead densities, they contain MA components, whose structure is unexploited when forming density forecasts. Therefore the forecast errors have an MA structure, just like multi-step-ahead forecast errors obtained from linear models.

Next, we move on to consider processes which allow for richer and more complex behavior. These models are more naturally estimated in recursive windows rather rolling windows. The latter scheme, by limiting the dependence on a finite lags of observables, might entail a loss of parameter estimation efficiency.

DGP S6: Random walk with jumps. This is a discrete version of the Merton (1976) jump diffusion process, given by

$$S_{t+1} = S_t \cdot \exp\left(\left(\mu - \frac{1}{2}\sigma^2 \right) + \sigma\epsilon_{t+1} + \sum_{i=1}^{N_{t+1}} y_{i,t+1} \right)\,, \tag{21}$$

where $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$, $y_{i,t} \overset{iid}{\sim} \mathcal{N}(-0.5\delta^2, \delta^2)$ and $N_t \overset{iid}{\sim} \mathscr{P}(\lambda)$, where $\mathscr{P}(\lambda)$ denotes the Poisson distribution with parameter λ, and S_t stands for the stock price at time t. Our variable of interest is the cumulative h-period logarithmic return given by $y_{t+h} \equiv \ln\left(\frac{S_{t+h}}{S_t}\right)$. The predictive CDF reads

$$\Phi_0(y_{t+h}|\mathscr{F}_t) = \sum_{n=0}^{\infty} F_Z\left(\frac{y_{t+h} - a_n}{s_n}\right) e^{-\lambda h} \frac{(\lambda h)^n}{n!}, \tag{22}$$

where $F_Z(\cdot)$ is the standard normal CDF, $a_n = (\mu - 0.5\sigma^2)h - 0.5\delta^2 n$ and $s_n^2 = h\sigma^2 + n\delta^2$.[9] Parameter values, based on S&P 500 index returns expressed in percentage form on a daily basis, are taken from Table VI (Column BSJ) in Andersen et al. (2002): $\mu = 0.0631, \sigma = 0.7166$, $\delta = 0.0135$ and $\lambda = 0.0591$. Such value of λ implies approximately 15 jumps per business year. We consider $h = \{1, 5\}$, corresponding to one-day-ahead and one-week-ahead returns, respectively.

DGP S7: GARCH(1,1). Despite many variations and developments, the Gaussian GARCH(1,1) model remains competitive among variants of the same family (see *e.g.*, Hansen and Lunde, 2005) or against newer forecasting approaches such as quantile regressions in the Growth-at-Risk literature (see *e.g.*, Brownlees and Souza, 2021). The GARCH(1,1) model is given by

$$y_t = \sigma_t \varepsilon_t, \tag{23}$$

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{24}$$

where $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$. We set the parameter values as $\omega = 0.0070$, $\alpha = 0.1314$ and $\beta = 0.8563$. These are provided by Volatility Laboratory (V-Lab) at NYU Stern and correspond to GARCH(1,1) parameters estimated using the Mexican Peso-USD exchange rate from May 31, 1995 to June 6, 2025.[10] Notice that $\alpha + \beta \approx 1$, indicating that near unit-root behavior is present in the volatility equation, as is common in exchange rate series. For y_t being daily log returns, y_{t+h} corresponds to the daily return h days ahead. *i.e.*, the log return of holding the asset from day $t + h - 1$ to $t + h$. Therefore the results of DGP S6 and S7 are not directly comparable as different objects are forecast. In the exercises, we consider forecasts both one and five steps ahead, corresponding to one-day-ahead and one-week-ahead forecasts. For $h = 1$ the predictive CDF is normal: $\Phi_0(y_{t+1}|\mathscr{F}_t) = \mathcal{N}(0, \omega + \alpha y_t^2 + \beta \sigma_t^2)$, while for $h = 5$ we approximate it by simulating the model forward 5000 times.

[9]In the simulations, the infinite sum in Equation (22) is truncated at the 99.99th percentile of the compound Poisson process.

[10]The data and parameter estimates are available on the V-Lab website at https://vlab.stern.nyu.edu/.

Size results for DGPs S1–S3 are shown in Table 2. For sample sizes up to and including $P = 200$, the finite-sample critical values tabulated in Table 1 are used, while results for $P = \{500, 1000\}$ are based on the tabulated asymptotic critical values. The table shows that the tests maintain correct size in both small and large samples. In line with the results reported in Rossi and Sekhposyan (2019), the use of finite-sample critical values leads to improved size properties in small samples, as shown in Appendix C.

Table 3 shows the empirical rejection rates of the tests for multi-step-ahead DGPs, S4 and S5. In the case of DGP S4, Panel A demonstrates that most configurations lead to desirable size control even for small sample sizes. As two exceptions, we see overrejection for small to moderate sample sizes (up to $P = 200$) when restricting the domain of r to either the left or the right tail, and underrejection when focusing on the tails using the weight function in conjunction with the Kolmogorov–Smirnov statistic. However, for large sample sizes all tests reject approximately at the nominal level of 5%. Turning to DGP S5, Panels B to D generally show reasonable size properties, albeit with empirical rejection rates gradually approaching the nominal 5% level from below when considering the full distribution, *i.e.*, $r \in [0, 1]$. Additionally, we observe a pattern of over- and underrejections when focusing on the tails, similar to the case of DGP S4.

All simulation results shown in Table 3 are based on setting the bootstrap block length as $\ell = \max\{h - 1, \lfloor P^{1/3} \rfloor\}$. While asymptotically, as $P \to \infty$ with finite and fixed h, this is equivalent to setting $\ell = \lfloor P^{1/3} \rfloor$, we see slightly better empirical rejection rates when the choice of block length takes into account the maximum dependence of the PITs under the null hypothesis, and hence we suggest that empirical researchers use this rule. See Appendix C for further details.

Table 4 shows the empirical rejection rates of DGPs S6 and S7 at horizons of $h = 1$ and $h = 5$. At the one-step-ahead horizon (Panels A and C), all tests show excellent size control even with $P = 250$ out-of-sample observations, which corresponds to approximately one business year. At the five-step-ahead horizon, Panels B and D show some overrejection in small to moderately sized samples (up to $P = 1000$) especially when restricting attention to either the left or the right tail ($r \in [0, 0.1]$ or $r \in [0.9, 1]$, respectively), while the weighted counterparts ($w(r) = (1 - r)^2$ or $w(r) = r^2$, respectively) are somewhat less affected. Nevertheless, as the sample size reaches $P = 2500$, corresponding to approximately ten business years, the empirical rejection rates are in the close vicinity of the nominal size of 5%.

Table 2: Size properties: DGPs S1–S3

Panel A: DGP S1

Region or weight function		$P = 25$		$P = 50$		$P = 100$		$P = 200$		$P = 500$		$P = 1000$	
		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0, 1]$	0.049	0.051	0.049	0.049	0.051	0.047	0.044	0.046	0.054	0.051	0.048	0.050
Left tail	$r \in [0, 0.1]$	0.051	0.049	0.047	0.047	0.048	0.051	0.051	0.052	0.054	0.049	0.044	0.050
Right tail	$r \in [0.9, 1]$	0.051	0.051	0.050	0.049	0.048	0.048	0.051	0.051	0.051	0.048	0.042	0.050
Left half	$r \in [0, 0.5]$	0.051	0.052	0.050	0.047	0.049	0.047	0.046	0.047	0.049	0.053	0.051	0.051
Right half	$r \in [0.5, 1]$	0.051	0.051	0.049	0.050	0.050	0.047	0.045	0.045	0.048	0.049	0.051	0.050
Center	$r \in [0.1, 0.9]$	0.048	0.052	0.049	0.050	0.051	0.048	0.044	0.046	0.054	0.050	0.048	0.049
Tails	$r \in \{[0, 0.1] \cup [0.9, 1]\}$	0.051	0.050	0.050	0.051	0.046	0.050	0.050	0.048	0.045	0.050	0.049	0.051
Left tail	$w(r) = (1 - r)^2$	0.049	0.052	0.049	0.048	0.049	0.048	0.053	0.047	0.050	0.052	0.051	0.051
Right tail	$w(r) = r^2$	0.050	0.051	0.048	0.048	0.050	0.048	0.048	0.047	0.045	0.050	0.048	0.050
Center	$w(r) = r(1 - r)$	0.050	0.052	0.049	0.049	0.051	0.049	0.046	0.046	0.051	0.050	0.049	0.048
Tails	$w(r) = (2r - 1)^2$	0.051	0.051	0.049	0.046	0.046	0.050	0.051	0.047	0.050	0.049	0.052	0.052

Panel B: DGP S2

Region or weight function		$P = 25$		$P = 50$		$P = 100$		$P = 200$		$P = 500$		$P = 1000$	
		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0, 1]$	0.054	0.051	0.049	0.047	0.047	0.049	0.050	0.051	0.052	0.053	0.048	0.046
Left tail	$r \in [0, 0.1]$	0.046	0.046	0.054	0.055	0.048	0.050	0.049	0.048	0.051	0.048	0.041	0.047
Right tail	$r \in [0.9, 1]$	0.050	0.048	0.051	0.050	0.049	0.049	0.049	0.050	0.051	0.048	0.043	0.049
Left half	$r \in [0, 0.5]$	0.049	0.051	0.051	0.051	0.048	0.047	0.049	0.050	0.046	0.049	0.051	0.047
Right half	$r \in [0.5, 1]$	0.055	0.054	0.048	0.047	0.050	0.052	0.051	0.054	0.048	0.052	0.052	0.051
Center	$r \in [0.1, 0.9]$	0.053	0.051	0.049	0.046	0.047	0.051	0.050	0.052	0.052	0.053	0.048	0.046
Tails	$r \in \{[0, 0.1] \cup [0.9, 1]\}$	0.048	0.047	0.053	0.054	0.048	0.048	0.048	0.049	0.045	0.049	0.049	0.049
Left tail	$w(r) = (1 - r)^2$	0.047	0.051	0.051	0.051	0.048	0.047	0.048	0.049	0.048	0.051	0.045	0.046
Right tail	$w(r) = r^2$	0.052	0.053	0.051	0.048	0.052	0.052	0.054	0.053	0.050	0.052	0.051	0.051
Center	$w(r) = r(1 - r)$	0.054	0.052	0.049	0.048	0.049	0.051	0.049	0.052	0.049	0.053	0.051	0.047
Tails	$w(r) = (2r - 1)^2$	0.047	0.049	0.054	0.051	0.050	0.047	0.049	0.049	0.050	0.048	0.048	0.050

Panel C: DGP S3

Region or weight function		$P = 25$		$P = 50$		$P = 100$		$P = 200$		$P = 500$		$P = 1000$	
		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0, 1]$	0.049	0.049	0.045	0.046	0.050	0.048	0.045	0.048	0.053	0.051	0.045	0.047
Left tail	$r \in [0, 0.1]$	0.045	0.044	0.046	0.047	0.049	0.049	0.050	0.051	0.052	0.049	0.039	0.047
Right tail	$r \in [0.9, 1]$	0.050	0.048	0.048	0.048	0.051	0.051	0.050	0.047	0.050	0.048	0.042	0.053
Left half	$r \in [0, 0.5]$	0.047	0.049	0.046	0.044	0.049	0.046	0.048	0.047	0.049	0.053	0.048	0.047
Right half	$r \in [0.5, 1]$	0.051	0.051	0.049	0.045	0.048	0.048	0.048	0.050	0.047	0.049	0.049	0.047
Center	$r \in [0.1, 0.9]$	0.049	0.049	0.045	0.046	0.050	0.048	0.045	0.048	0.053	0.052	0.045	0.048
Tails	$r \in \{[0, 0.1] \cup [0.9, 1]\}$	0.049	0.047	0.045	0.047	0.049	0.048	0.049	0.048	0.045	0.050	0.043	0.049
Left tail	$w(r) = (1 - r)^2$	0.046	0.050	0.047	0.046	0.049	0.045	0.047	0.049	0.049	0.053	0.048	0.045
Right tail	$w(r) = r^2$	0.051	0.052	0.050	0.047	0.049	0.047	0.049	0.049	0.048	0.049	0.049	0.049
Center	$w(r) = r(1 - r)$	0.048	0.049	0.049	0.044	0.049	0.047	0.048	0.048	0.052	0.052	0.049	0.048
Tails	$w(r) = (2r - 1)^2$	0.046	0.052	0.046	0.048	0.048	0.048	0.049	0.046	0.050	0.049	0.047	0.046

Note: The table reports the empirical rejection rates of the Kolmogorov–Smirnov- (KS) and Cramér–von Mises- (CvM) type tests in various regions and with various weights (in rows) for different sample sizes, denoted by P (in columns) at the 5% nominal size for DGPs S1–S3. For sample sizes up to and including $P = 200$, the finite-sample critical values reported in Table 1 are used, while results for $P = 500$ and $P = 1000$ are based on the tabulated asymptotic critical values. The number of Monte Carlo simulations is 10,000. The domain of r is discretized over $[0 : 0.001 : 1]$.

Table 3: Size properties: DGPs S4 and S5

Panel A: DGP S4

Region or weight function		$P=25$		$P=50$		$P=100$		$P=200$		$P=500$		$P=1000$	
		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0,1]$	0.055	0.057	0.055	0.053	0.061	0.058	0.055	0.056	0.054	0.054	0.053	0.054
Left tail	$r \in [0,0.1]$	0.127	0.162	0.109	0.104	0.089	0.080	0.076	0.071	0.057	0.056	0.058	0.055
Right tail	$r \in [0.9,1]$	0.131	0.167	0.114	0.111	0.090	0.081	0.075	0.069	0.062	0.062	0.057	0.057
Left half	$r \in [0,0.5]$	0.068	0.064	0.061	0.059	0.063	0.059	0.059	0.057	0.057	0.055	0.052	0.054
Right half	$r \in [0.5,1]$	0.067	0.066	0.060	0.060	0.063	0.056	0.056	0.055	0.054	0.051	0.052	0.052
Center	$r \in [0.1,0.9]$	0.056	0.057	0.055	0.053	0.061	0.058	0.055	0.056	0.054	0.054	0.053	0.053
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.047	0.059	0.050	0.053	0.048	0.053	0.056	0.055	0.051	0.053	0.054	0.052
Left tail	$w(r) = (1-r)^2$	0.073	0.064	0.070	0.057	0.064	0.061	0.061	0.056	0.056	0.055	0.052	0.054
Right tail	$w(r) = r^2$	0.075	0.065	0.067	0.060	0.063	0.057	0.057	0.054	0.055	0.053	0.055	0.052
Center	$w(r) = r(1-r)$	0.066	0.058	0.061	0.054	0.062	0.057	0.057	0.056	0.057	0.054	0.053	0.053
Tails	$w(r) = (2r-1)^2$	0.011	0.049	0.023	0.053	0.031	0.057	0.044	0.054	0.050	0.055	0.051	0.054

Panel B: DGP S5, $h = 2$

Region or weight function		$P=25$		$P=50$		$P=100$		$P=200$		$P=500$		$P=1000$	
		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0,1]$	0.025	0.018	0.035	0.026	0.036	0.027	0.041	0.031	0.045	0.038	0.043	0.043
Left tail	$r \in [0,0.1]$	0.094	0.127	0.085	0.086	0.070	0.062	0.061	0.055	0.053	0.049	0.051	0.051
Right tail	$r \in [0.9,1]$	0.099	0.134	0.083	0.080	0.070	0.061	0.062	0.058	0.052	0.050	0.054	0.050
Left half	$r \in [0,0.5]$	0.029	0.022	0.038	0.030	0.041	0.033	0.042	0.037	0.046	0.042	0.043	0.043
Right half	$r \in [0.5,1]$	0.032	0.025	0.036	0.029	0.039	0.030	0.042	0.034	0.043	0.038	0.043	0.043
Center	$r \in [0.1,0.9]$	0.025	0.018	0.035	0.026	0.036	0.028	0.041	0.031	0.045	0.037	0.043	0.042
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.049	0.062	0.054	0.055	0.050	0.050	0.052	0.049	0.048	0.048	0.051	0.047
Left tail	$w(r) = (1-r)^2$	0.037	0.021	0.044	0.030	0.046	0.032	0.050	0.036	0.047	0.042	0.051	0.044
Right tail	$w(r) = r^2$	0.040	0.024	0.047	0.031	0.042	0.029	0.048	0.033	0.046	0.035	0.049	0.045
Center	$w(r) = r(1-r)$	0.029	0.019	0.035	0.025	0.038	0.028	0.041	0.032	0.045	0.038	0.042	0.042
Tails	$w(r) = (2r-1)^2$	0.013	0.027	0.028	0.032	0.036	0.034	0.044	0.038	0.043	0.039	0.051	0.045

Panel C: DGP S5, $h = 4$

Region or weight function		$P=25$		$P=50$		$P=100$		$P=200$		$P=500$		$P=1000$	
		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0,1]$	0.019	0.010	0.024	0.015	0.029	0.018	0.034	0.026	0.035	0.028	0.040	0.036
Left tail	$r \in [0,0.1]$	0.085	0.118	0.081	0.079	0.061	0.057	0.057	0.050	0.050	0.048	0.050	0.047
Right tail	$r \in [0.9,1]$	0.090	0.122	0.076	0.074	0.066	0.056	0.056	0.049	0.049	0.045	0.054	0.050
Left half	$r \in [0,0.5]$	0.025	0.017	0.029	0.020	0.032	0.025	0.038	0.031	0.035	0.031	0.040	0.038
Right half	$r \in [0.5,1]$	0.023	0.016	0.028	0.020	0.031	0.025	0.036	0.032	0.036	0.032	0.040	0.037
Center	$r \in [0.1,0.9]$	0.019	0.010	0.024	0.015	0.029	0.018	0.034	0.026	0.035	0.028	0.040	0.037
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.052	0.060	0.051	0.050	0.043	0.044	0.047	0.042	0.044	0.045	0.050	0.050
Left tail	$w(r) = (1-r)^2$	0.031	0.016	0.039	0.019	0.040	0.024	0.043	0.029	0.039	0.031	0.043	0.039
Right tail	$w(r) = r^2$	0.034	0.017	0.038	0.019	0.040	0.023	0.042	0.029	0.041	0.032	0.044	0.037
Center	$w(r) = r(1-r)$	0.023	0.011	0.026	0.016	0.031	0.018	0.035	0.027	0.036	0.029	0.042	0.037
Tails	$w(r) = (2r-1)^2$	0.016	0.021	0.025	0.025	0.030	0.026	0.039	0.032	0.040	0.035	0.045	0.040

Panel D: DGP S5, $h = 12$

Region or weight function		$P=25$		$P=50$		$P=100$		$P=200$		$P=500$		$P=1000$	
		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0,1]$	0.049	0.052	0.023	0.021	0.031	0.026	0.038	0.028	0.039	0.036	0.038	0.035
Left tail	$r \in [0,0.1]$	0.070	0.088	0.068	0.061	0.057	0.054	0.060	0.057	0.053	0.051	0.048	0.050
Right tail	$r \in [0.9,1]$	0.068	0.083	0.070	0.062	0.067	0.060	0.060	0.056	0.049	0.047	0.045	0.048
Left half	$r \in [0,0.5]$	0.065	0.071	0.028	0.027	0.035	0.031	0.041	0.035	0.041	0.037	0.041	0.038
Right half	$r \in [0.5,1]$	0.066	0.068	0.031	0.029	0.036	0.030	0.040	0.035	0.044	0.040	0.039	0.038
Center	$r \in [0.1,0.9]$	0.050	0.057	0.024	0.023	0.032	0.027	0.038	0.028	0.039	0.037	0.038	0.035
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.056	0.056	0.052	0.046	0.049	0.047	0.054	0.050	0.049	0.045	0.044	0.046
Left tail	$w(r) = (1-r)^2$	0.058	0.058	0.038	0.025	0.038	0.029	0.045	0.033	0.047	0.039	0.045	0.038
Right tail	$w(r) = r^2$	0.061	0.057	0.039	0.027	0.041	0.029	0.042	0.034	0.048	0.039	0.042	0.036
Center	$w(r) = r(1-r)$	0.066	0.059	0.030	0.024	0.035	0.028	0.041	0.029	0.039	0.037	0.040	0.036
Tails	$w(r) = (2r-1)^2$	0.043	0.045	0.035	0.023	0.038	0.032	0.045	0.037	0.043	0.040	0.041	0.038

Note: The table reports the empirical rejection rates of the Kolmogorov–Smirnov- (KS) and Cramér–von Mises- (CvM) type tests in various regions and with various weights (in rows) for different sample sizes, denoted by P (in columns) at the 5% nominal size for DGPs S4 and S5. For all sample sizes P, the critical values are based on 5000 bootstrap replications with block length $\ell = \max\{h-1, \lfloor P^{1/3} \rfloor\}$. The number of Monte Carlo simulations is 10,000. The domain of r is discretized over $[0 : 0.001 : 1]$.

Table 4: Size properties: DGPs S6 and S7

Panel A: DGP S6, $h = 1$

Region or weight function		$P = 250$ KS	$P = 250$ CvM	$P = 500$ KS	$P = 500$ CvM	$P = 1000$ KS	$P = 1000$ CvM	$P = 2500$ KS	$P = 2500$ CvM
Full	$r \in [0,1]$	0.050	0.048	0.057	0.052	0.049	0.048	0.054	0.055
Left tail	$r \in [0,0.1]$	0.047	0.049	0.055	0.053	0.041	0.048	0.053	0.052
Right tail	$r \in [0.9,1]$	0.046	0.052	0.056	0.054	0.043	0.046	0.050	0.052
Left half	$r \in [0,0.5]$	0.050	0.051	0.051	0.052	0.051	0.051	0.051	0.055
Right half	$r \in [0.5,1]$	0.049	0.050	0.053	0.053	0.051	0.049	0.051	0.052
Center	$r \in [0.1,0.9]$	0.050	0.049	0.057	0.052	0.049	0.048	0.054	0.055
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.055	0.051	0.051	0.054	0.048	0.048	0.057	0.053
Left tail	$w(r) = (1-r)^2$	0.049	0.049	0.051	0.054	0.049	0.050	0.053	0.055
Right tail	$w(r) = r^2$	0.051	0.051	0.055	0.055	0.049	0.049	0.052	0.052
Center	$w(r) = r(1-r)$	0.050	0.049	0.054	0.051	0.050	0.049	0.051	0.054
Tails	$w(r) = (2r-1)^2$	0.054	0.050	0.053	0.053	0.050	0.049	0.053	0.053

Panel B: DGP S6, $h = 5$

Region or weight function		$P = 250$ KS	$P = 250$ CvM	$P = 500$ KS	$P = 500$ CvM	$P = 1000$ KS	$P = 1000$ CvM	$P = 2500$ KS	$P = 2500$ CvM
Full	$r \in [0,1]$	0.090	0.093	0.084	0.086	0.074	0.077	0.065	0.065
Left tail	$r \in [0,0.1]$	0.117	0.121	0.096	0.098	0.076	0.081	0.068	0.071
Right tail	$r \in [0.9,1]$	0.116	0.125	0.094	0.097	0.083	0.089	0.067	0.067
Left half	$r \in [0,0.5]$	0.089	0.088	0.086	0.087	0.071	0.074	0.063	0.065
Right half	$r \in [0.5,1]$	0.093	0.093	0.080	0.083	0.078	0.077	0.065	0.064
Center	$r \in [0.1,0.9]$	0.090	0.093	0.084	0.086	0.074	0.076	0.065	0.065
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.071	0.091	0.073	0.086	0.068	0.078	0.065	0.067
Left tail	$w(r) = (1-r)^2$	0.094	0.091	0.089	0.087	0.075	0.077	0.069	0.065
Right tail	$w(r) = r^2$	0.098	0.094	0.083	0.086	0.076	0.078	0.064	0.065
Center	$w(r) = r(1-r)$	0.090	0.091	0.084	0.084	0.074	0.076	0.065	0.065
Tails	$w(r) = (2r-1)^2$	0.060	0.097	0.063	0.087	0.066	0.079	0.063	0.066

Panel C: DGP S7, $h = 1$

Region or weight function		$P = 250$ KS	$P = 250$ CvM	$P = 500$ KS	$P = 500$ CvM	$P = 1000$ KS	$P = 1000$ CvM	$P = 2500$ KS	$P = 2500$ CvM
Full	$r \in [0,1]$	0.049	0.048	0.050	0.048	0.049	0.053	0.052	0.051
Left tail	$r \in [0,0.1]$	0.049	0.053	0.054	0.051	0.041	0.047	0.049	0.047
Right tail	$r \in [0.9,1]$	0.045	0.046	0.049	0.047	0.041	0.048	0.046	0.046
Left half	$r \in [0,0.5]$	0.050	0.053	0.046	0.050	0.051	0.053	0.055	0.052
Right half	$r \in [0.5,1]$	0.047	0.048	0.048	0.049	0.052	0.050	0.047	0.048
Center	$r \in [0.1,0.9]$	0.049	0.048	0.050	0.048	0.049	0.053	0.052	0.051
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.051	0.048	0.044	0.049	0.048	0.049	0.050	0.046
Left tail	$w(r) = (1-r)^2$	0.055	0.053	0.050	0.049	0.051	0.052	0.053	0.054
Right tail	$w(r) = r^2$	0.044	0.048	0.048	0.050	0.051	0.052	0.047	0.050
Center	$w(r) = r(1-r)$	0.047	0.048	0.047	0.048	0.051	0.053	0.051	0.051
Tails	$w(r) = (2r-1)^2$	0.056	0.051	0.049	0.051	0.052	0.052	0.046	0.051

Panel D: DGP S7, $h = 5$

Region or weight function		$P = 250$ KS	$P = 250$ CvM	$P = 500$ KS	$P = 500$ CvM	$P = 1000$ KS	$P = 1000$ CvM	$P = 2500$ KS	$P = 2500$ CvM
Full	$r \in [0,1]$	0.059	0.052	0.054	0.051	0.053	0.053	0.053	0.050
Left tail	$r \in [0,0.1]$	0.078	0.079	0.072	0.068	0.059	0.060	0.057	0.055
Right tail	$r \in [0.9,1]$	0.080	0.078	0.070	0.067	0.063	0.063	0.058	0.056
Left half	$r \in [0,0.5]$	0.059	0.056	0.052	0.054	0.054	0.054	0.052	0.052
Right half	$r \in [0.5,1]$	0.057	0.054	0.056	0.053	0.051	0.051	0.051	0.052
Center	$r \in [0.1,0.9]$	0.059	0.052	0.054	0.051	0.052	0.052	0.053	0.050
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.066	0.070	0.066	0.067	0.058	0.059	0.056	0.056
Left tail	$w(r) = (1-r)^2$	0.067	0.056	0.063	0.057	0.055	0.052	0.055	0.052
Right tail	$w(r) = r^2$	0.067	0.055	0.061	0.054	0.055	0.052	0.056	0.051
Center	$w(r) = r(1-r)$	0.057	0.052	0.053	0.051	0.053	0.052	0.052	0.050
Tails	$w(r) = (2r-1)^2$	0.058	0.062	0.057	0.064	0.054	0.059	0.054	0.055

Note: The table reports the empirical rejection rates of the Kolmogorov–Smirnov- (KS) and Cramér–von Mises- (CvM) type tests in various regions and with various weights (in rows) for different sample sizes, denoted by P (in columns) at the 5% nominal size for DGPs S6 and S7, at horizons $h = 1$ and $h = 5$. For $h = 1$ the asymptotic critical values reported in Table 1 are used, while for $h = 5$, the critical values are based on 5000 bootstrap simulations. The number of Monte Carlo simulations is 10,000. The domain of r is discretized over $[0 : 0.001 : 1]$.

4.2 Power analysis

In the first set of power analysis exercises (DGPs P1-P4) we modify the innovation term ε_t of the baseline model (DGP S1) and the serial correlation model (DGP S4) in two different directions, following the spirit of the simulation examples in Rossi and Sekhposyan (2019). In these simulations, the sample size is $P = 1000$.

DGPs P1 and P2: Linear combination of normal and chi-squared. In the first exercises, DGPs P1 and P2, derived from DGPs S1 and S4, respectively, we set the error term ε_t as a linear combination of a standard normal and a rescaled chi-squared with one degree of freedom: $\varepsilon_t = \sqrt{1-c} \cdot \eta_{1,t} + \sqrt{c} \cdot (\eta_{2,t}^2 - 1)/\sqrt{2}$, where $\eta_{i,t} \stackrel{iid}{\sim} \mathcal{N}(0,1)$ for $i = 1,2$, and $\eta_{1,t}$ independent of $\eta_{2,s}$ for all t and s. In this setup, $c = 0$ corresponds to the null, and by increasing its value up to $c = 0.6$ we deviate further and further from the null. Notice that ε_t has expected value zero and unit variance for any $0 \leq c \leq 1$.

DGPs P3 and P4: Student-*t*. In the second set of exercises, labeled DGPs P3 and P4, we assume that the innovation term in the baseline and the serial correlation models, respectively, follows a Student-t distribution with ν degrees of freedom, appropriately rescaled to have unit variance: $\varepsilon_t \stackrel{iid}{\sim} \sqrt{\frac{\nu-2}{\nu}} t_\nu$. When $\nu \to \infty$, we are under the null, and smaller values of ν correspond to larger deviations from the null.

Next, we investigate the power properties of the tests in a non-linear, non-Gaussian setup.

DGP P5: Random walk without jumps. In this exercise, the true DGP under the alternative is based on DGP S6, but under the null we use a misspecified normal predictive CDF, where the presence of jumps is ignored, corresponding to $\lambda = 0$ (or $\delta = 0$). As the frequency λ and size δ of the jumps increases, misspecification also increases. Since in these models, especially in the absence of stochastic volatility, the magnitude of the jumps is very small, for the power exercise we set the size of the jump to three times as large as the standard deviation of the diffusion component, $\delta = 3\sigma = 3 \cdot 0.7166$, and use λ to regulate the deviation from the null. Based on the size properties reported in Table 4, we set $P = 2500$ in the power exercises.

For DGPs P1 and P2, the results in Table 5 indicate that in general, both Kolmogorov–Smirnov- and Cramér–von Mises-type tests have good power in detecting misspecification in predictive densities at both $h = 1$ and $h = 2$.[11] As an exception, however, when restricting attention to the right tail, *i.e.*, $r \in [0.9, 1]$, the tests fail to achieve satisfactory power even when $c = 0.60$. This is due to the fact that the linear combination of the normal and the chi-squared distributed

[11] Additional results reported in Appendix C indicate that the tests still have power when considering a smaller sample size, $P = 100$.

innovations, when viewed through the lens of the normal CDF, is difficult to distinguish from a normally distributed innovation's upper tail. To illustrate this problem, Panel (a) of Figure 1 shows the CDFs of the PITs when the innovations are distributed according to the linear combination, but evaluated through the normal CDF. As we can see, deviations in the right tail are rather small. Alternatively, one could use the weighting function $w(r) = r^2$, which puts emphasis on the right tail of the distribution without completely discarding deviations from uniformity in other regions. As the results in Table 5 confirm, this strategy can lead to better power.

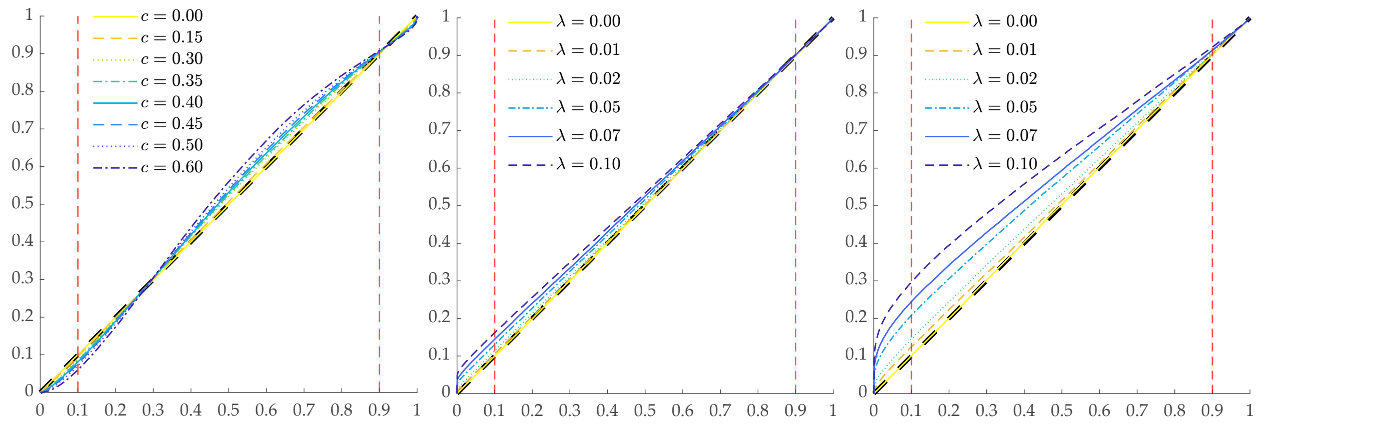
Table 5: Power properties: DGPs P1 and P2

Panel A: DGP P1, $h = 1$ (linear combination of normal and chi-squared)																	
		$c = 0.00$		$c = 0.15$		$c = 0.30$		$c = 0.35$		$c = 0.40$		$c = 0.45$		$c = 0.50$		$c = 0.60$	
Region or weight function		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0,1]$	0.05	0.05	0.08	0.07	0.32	0.31	0.47	0.49	0.65	0.68	0.82	0.87	0.94	0.97	1.00	1.00
Left tail	$r \in [0,0.1]$	0.05	0.05	0.06	0.08	0.28	0.40	0.45	0.60	0.66	0.79	0.84	0.94	0.96	0.99	1.00	1.00
Right tail	$r \in [0.9,1]$	0.04	0.05	0.05	0.06	0.08	0.11	0.10	0.13	0.15	0.17	0.22	0.24	0.29	0.29	0.50	0.50
Left half	$r \in [0,0.5]$	0.05	0.05	0.07	0.06	0.21	0.12	0.31	0.19	0.45	0.31	0.63	0.52	0.82	0.77	1.00	1.00
Right half	$r \in [0.5,1]$	0.05	0.05	0.08	0.09	0.37	0.39	0.53	0.56	0.70	0.73	0.85	0.87	0.95	0.96	1.00	1.00
Center	$r \in [0.1,0.9]$	0.05	0.05	0.08	0.07	0.32	0.30	0.47	0.46	0.65	0.65	0.82	0.83	0.94	0.95	1.00	1.00
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.05	0.05	0.06	0.08	0.23	0.37	0.38	0.57	0.58	0.79	0.79	0.94	0.94	0.99	1.00	1.00
Left tail	$w(r) = (1-r)^2$	0.05	0.05	0.06	0.06	0.21	0.20	0.34	0.35	0.54	0.58	0.75	0.83	0.92	0.97	1.00	1.00
Right tail	$w(r) = r^2$	0.05	0.05	0.07	0.08	0.27	0.36	0.42	0.53	0.60	0.70	0.78	0.86	0.91	0.96	1.00	1.00
Center	$w(r) = r(1-r)$	0.05	0.05	0.08	0.08	0.31	0.31	0.46	0.46	0.62	0.64	0.79	0.82	0.91	0.94	0.99	1.00
Tails	$w(r) = (2r-1)^2$	0.05	0.05	0.07	0.06	0.31	0.31	0.50	0.54	0.72	0.79	0.90	0.95	0.98	1.00	1.00	1.00
Panel B: DGP P2, $h = 2$ (linear combination of normal and chi-squared)																	
		$c = 0.00$		$c = 0.15$		$c = 0.30$		$c = 0.35$		$c = 0.40$		$c = 0.45$		$c = 0.50$		$c = 0.60$	
Region or weight function		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0,1]$	0.05	0.05	0.07	0.07	0.25	0.24	0.37	0.36	0.52	0.53	0.68	0.71	0.84	0.88	0.99	1.00
Left tail	$r \in [0,0.1]$	0.05	0.06	0.09	0.11	0.32	0.42	0.47	0.59	0.66	0.78	0.82	0.91	0.94	0.98	1.00	1.00
Right tail	$r \in [0.9,1]$	0.06	0.05	0.06	0.06	0.09	0.07	0.12	0.08	0.16	0.10	0.22	0.13	0.31	0.16	0.52	0.28
Left half	$r \in [0,0.5]$	0.05	0.05	0.07	0.06	0.16	0.11	0.24	0.15	0.34	0.23	0.46	0.37	0.64	0.58	0.95	0.96
Right half	$r \in [0.5,1]$	0.05	0.05	0.08	0.08	0.30	0.31	0.43	0.45	0.58	0.60	0.72	0.74	0.85	0.87	0.98	0.99
Center	$r \in [0.1,0.9]$	0.05	0.06	0.07	0.07	0.25	0.23	0.37	0.35	0.52	0.50	0.68	0.67	0.84	0.84	0.99	0.99
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.05	0.05	0.07	0.09	0.23	0.33	0.37	0.51	0.55	0.71	0.75	0.88	0.90	0.97	1.00	1.00
Left tail	$w(r) = (1-r)^2$	0.05	0.05	0.07	0.06	0.19	0.15	0.28	0.26	0.44	0.42	0.65	0.66	0.83	0.87	1.00	1.00
Right tail	$w(r) = r^2$	0.05	0.05	0.08	0.08	0.26	0.28	0.38	0.41	0.54	0.57	0.69	0.71	0.84	0.85	0.99	0.98
Center	$w(r) = r(1-r)$	0.05	0.06	0.07	0.07	0.24	0.23	0.36	0.35	0.49	0.50	0.64	0.65	0.79	0.82	0.96	0.99
Tails	$w(r) = (2r-1)^2$	0.05	0.05	0.08	0.07	0.28	0.25	0.46	0.42	0.66	0.64	0.85	0.85	0.96	0.97	1.00	1.00

Note: The table reports the empirical rejection rates of the Kolmogorov–Smirnov- (KS) and Cramér–von Mises- (CvM) type tests in various regions and with various weights (in rows) for different degrees of deviations from the null, denoted by c (in columns, with $c = 0$ corresponding to the null) at the 5% nominal size for DGPs P1 and P2. The sample size is $P = 1000$ in all simulations, and for DGP P1 the asymptotic critical values reported in Table 1 are used, while for DGP P2, the critical values are based on 5000 bootstrap simulations. The number of Monte Carlo simulations is 10,000. The domain of r is discretized over $[0 : 0.001 : 1]$.

Table 6 shows the power properties of the tests for DGPs P3 and P4. As we can see, power is inversely related to the degrees of freedom ν. In general, the weighted versions of the tests focusing on the tails are more powerful than their counterparts restricting attention to the tails, but we observe the opposite when focusing on the center.

Figure 1: Cumulative distribution functions of PITs: DGPs P1 and P5



(a) DGP P1 (b) DGP P5, $h = 1$ (c) DGP P5, $h = 5$

Note: The panels show the empirical CDFs of the PITs of DGPs P1 and P5. The predictions are produced based on the alternative models (for various degrees of misspecification, indexed by c or λ) but the PITs are evaluated through the true predictive CDF under the null (corresponding to $c = 0$ or $\lambda = 0$). The dashed (black) diagonal line corresponds to the standard uniform CDF, which coincides with the CDF of the PITs under the null. The dashed (red) vertical lines mark the tail regions of $r \in [0, 0.1]$ and $r \in [0.9, 1]$. The number of Monte Carlo simulations is 100,000 in each case.

Table 6: Power properties: DGPs P3 and P4

		$\nu \to \infty$		$\nu = 15$		$\nu = 10$		$\nu = 7$		$\nu = 6$		$\nu = 5$		$\nu = 4$		$\nu = 3$	
Panel A: DGP P3, $h = 1$ (Student-t)																	
Region or weight function		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0, 1]$	0.05	0.05	0.09	0.07	0.17	0.14	0.41	0.41	0.63	0.67	0.89	0.93	1.00	1.00	1.00	1.00
Left tail	$r \in [0, 0.1]$	0.04	0.05	0.06	0.06	0.10	0.09	0.22	0.17	0.33	0.25	0.50	0.39	0.82	0.71	1.00	0.99
Right tail	$r \in [0.9, 1]$	0.04	0.05	0.06	0.06	0.11	0.09	0.22	0.17	0.32	0.24	0.51	0.40	0.81	0.70	1.00	0.99
Left half	$r \in [0, 0.5]$	0.05	0.05	0.08	0.09	0.14	0.15	0.31	0.32	0.46	0.47	0.70	0.71	0.96	0.96	1.00	1.00
Right half	$r \in [0.5, 1]$	0.05	0.05	0.08	0.09	0.14	0.16	0.30	0.32	0.46	0.47	0.71	0.71	0.96	0.96	1.00	1.00
Center	$r \in [0.1, 0.9]$	0.05	0.05	0.09	0.07	0.17	0.14	0.41	0.40	0.63	0.66	0.89	0.93	1.00	1.00	1.00	1.00
Tails	$r \in \{[0, 0.1] \cup [0.9, 1]\}$	0.05	0.05	0.07	0.06	0.13	0.10	0.30	0.23	0.43	0.36	0.66	0.60	0.94	0.92	1.00	1.00
Left tail	$w(r) = (1 - r)^2$	0.05	0.05	0.10	0.09	0.20	0.15	0.43	0.38	0.61	0.57	0.82	0.85	0.99	1.00	1.00	1.00
Right tail	$w(r) = r^2$	0.05	0.05	0.11	0.09	0.21	0.17	0.43	0.38	0.61	0.58	0.83	0.85	0.98	1.00	1.00	1.00
Center	$w(r) = r(1 - r)$	0.05	0.05	0.07	0.07	0.11	0.11	0.23	0.31	0.38	0.55	0.67	0.88	0.98	1.00	1.00	1.00
Tails	$w(r) = (2r - 1)^2$	0.05	0.05	0.08	0.11	0.16	0.28	0.39	0.67	0.58	0.87	0.83	0.99	0.99	1.00	1.00	1.00
Panel B: DGP P4, $h = 2$ (Student-t)																	
		$\nu \to \infty$		$\nu = 15$		$\nu = 10$		$\nu = 7$		$\nu = 6$		$\nu = 5$		$\nu = 4$		$\nu = 3$	
Region or weight function		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0, 1]$	0.05	0.05	0.08	0.07	0.14	0.12	0.30	0.28	0.46	0.48	0.73	0.78	0.98	0.99	1.00	1.00
Left tail	$r \in [0, 0.1]$	0.06	0.05	0.10	0.08	0.15	0.11	0.27	0.19	0.37	0.25	0.54	0.39	0.81	0.65	1.00	0.98
Right tail	$r \in [0.9, 1]$	0.06	0.06	0.10	0.08	0.15	0.11	0.27	0.19	0.37	0.26	0.54	0.39	0.81	0.66	1.00	0.99
Left half	$r \in [0, 0.5]$	0.05	0.05	0.08	0.09	0.12	0.13	0.22	0.25	0.32	0.36	0.52	0.55	0.83	0.85	1.00	1.00
Right half	$r \in [0.5, 1]$	0.05	0.05	0.07	0.09	0.11	0.13	0.23	0.25	0.32	0.36	0.51	0.55	0.85	0.87	1.00	1.00
Center	$r \in [0.1, 0.9]$	0.05	0.05	0.08	0.07	0.14	0.11	0.30	0.27	0.46	0.46	0.73	0.77	0.98	0.99	1.00	1.00
Tails	$r \in \{[0, 0.1] \cup [0.9, 1]\}$	0.05	0.05	0.10	0.08	0.16	0.12	0.33	0.23	0.47	0.36	0.69	0.56	0.94	0.88	1.00	1.00
Left tail	$w(r) = (1 - r)^2$	0.06	0.05	0.12	0.09	0.20	0.14	0.39	0.28	0.54	0.43	0.74	0.68	0.95	0.96	1.00	1.00
Right tail	$w(r) = r^2$	0.05	0.05	0.11	0.09	0.20	0.14	0.39	0.28	0.53	0.43	0.74	0.68	0.95	0.96	1.00	1.00
Center	$w(r) = r(1 - r)$	0.05	0.05	0.07	0.07	0.10	0.10	0.17	0.21	0.26	0.37	0.44	0.67	0.85	0.98	1.00	1.00
Tails	$w(r) = (2r - 1)^2$	0.05	0.05	0.10	0.11	0.18	0.22	0.39	0.54	0.57	0.75	0.81	0.94	0.99	1.00	1.00	1.00

Note: The table reports the empirical rejection rates of the Kolmogorov–Smirnov- (KS) and Cramér–von Mises- (CvM) type tests in various regions and with various weights (in rows) for different degrees of deviations from the null, denoted by ν (in columns, with $\nu \to \infty$ corresponding to the null) at the 5% nominal size for DGPs P3 and P4. The sample size is $P = 1000$ in all simulations, and for DGP P3 the asymptotic critical values reported in Table 1 are used, while for DGP P4, the critical values are based on 5000 bootstrap simulations. The number of Monte Carlo simulations is 10,000. The domain of r is discretized over $[0 : 0.001 : 1]$.

Table 7 displays the power properties of the tests for DGP P5. In general, the tests have power against deviations from the null. However, there are three interesting exceptions: for $h = 1$ (in Panel A) when restricting attention to the right tail or the right half ($r \in [0.9, 1]$ or $r \in [0.5, 1]$), and for $h = 5$ (in Panel B) when restricting attention to the right tail. To explain this, recall that under the alternative, $\lambda > 0$, hence jumps do occur, and the expected value of a jump is negative and equals $-0.5\delta^2 = -0.5 \cdot (3 \cdot 0.7166)^2$. This means that under the alternative, deviations from normality (or equivalently, deviations from uniformity) are most prevalent in the *left* tail of the distribution. Panels (b) and (c) of Figure 1 illustrate this, and explain why tests restricting attention to deviations in the upper regions lack sufficient power.

Table 7: Power properties: DGP P5

Panel A: DGP P5, $h = 1$													
		$\lambda = 0.00$		$\lambda = 0.01$		$\lambda = 0.02$		$\lambda = 0.05$		$\lambda = 0.07$		$\lambda = 0.10$	
Region or weight function		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0, 1]$	0.05	0.05	0.07	0.08	0.18	0.21	0.98	0.93	1.00	1.00	1.00	1.00
Left tail	$r \in [0, 0.1]$	0.05	0.05	0.26	0.36	0.79	0.89	1.00	1.00	1.00	1.00	1.00	1.00
Right tail	$r \in [0.9, 1]$	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06
Left half	$r \in [0, 0.5]$	0.05	0.05	0.08	0.11	0.22	0.31	0.99	0.99	1.00	1.00	1.00	1.00
Right half	$r \in [0.5, 1]$	0.05	0.05	0.05	0.05	0.09	0.08	0.32	0.25	0.55	0.44	0.86	0.74
Center	$r \in [0.1, 0.9]$	0.05	0.05	0.07	0.08	0.18	0.18	0.89	0.80	1.00	0.98	1.00	1.00
Tails	$r \in \{[0, 0.1] \cup [0.9, 1]\}$	0.06	0.05	0.19	0.29	0.68	0.83	1.00	1.00	1.00	1.00	1.00	1.00
Left tail	$w(r) = (1-r)^2$	0.05	0.05	0.20	0.13	0.72	0.39	1.00	1.00	1.00	1.00	1.00	1.00
Right tail	$w(r) = r^2$	0.05	0.05	0.05	0.05	0.06	0.10	0.10	0.38	0.17	0.68	0.34	0.95
Center	$w(r) = r(1-r)$	0.05	0.05	0.06	0.07	0.13	0.17	0.57	0.77	0.88	0.97	1.00	1.00
Tails	$w(r) = (2r-1)^2$	0.05	0.05	0.26	0.14	0.88	0.49	1.00	1.00	1.00	1.00	1.00	1.00
Panel B: DGP P5, $h = 5$													
		$\lambda = 0.00$		$\lambda = 0.01$		$\lambda = 0.02$		$\lambda = 0.05$		$\lambda = 0.07$		$\lambda = 0.10$	
Region or weight function		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0, 1]$	0.07	0.07	0.20	0.21	0.69	0.62	1.00	1.00	1.00	1.00	1.00	1.00
Left tail	$r \in [0, 0.1]$	0.07	0.07	0.52	0.63	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00
Right tail	$r \in [0.9, 1]$	0.07	0.07	0.08	0.08	0.11	0.10	0.26	0.20	0.40	0.30	0.62	0.46
Left half	$r \in [0, 0.5]$	0.07	0.07	0.22	0.27	0.73	0.76	1.00	1.00	1.00	1.00	1.00	1.00
Right half	$r \in [0.5, 1]$	0.07	0.07	0.13	0.12	0.34	0.26	0.95	0.85	1.00	0.98	1.00	1.00
Center	$r \in [0.1, 0.9]$	0.07	0.07	0.20	0.19	0.66	0.55	1.00	1.00	1.00	1.00	1.00	1.00
Tails	$r \in \{[0, 0.1] \cup [0.9, 1]\}$	0.06	0.07	0.47	0.57	0.96	0.98	1.00	1.00	1.00	1.00	1.00	1.00
Left tail	$w(r) = (1-r)^2$	0.06	0.07	0.47	0.30	0.97	0.81	1.00	1.00	1.00	1.00	1.00	1.00
Right tail	$w(r) = r^2$	0.07	0.07	0.09	0.13	0.17	0.32	0.65	0.95	0.92	1.00	1.00	1.00
Center	$w(r) = r(1-r)$	0.07	0.07	0.16	0.18	0.46	0.53	1.00	1.00	1.00	1.00	1.00	1.00
Tails	$w(r) = (2r-1)^2$	0.06	0.07	0.57	0.34	0.99	0.87	1.00	1.00	1.00	1.00	1.00	1.00

Note: The table reports the empirical rejection rates of the Kolmogorov–Smirnov- (KS) and Cramér–von Mises- (CvM) type tests in various regions and with various weights (in rows) for different degrees of deviations from the null, denoted by λ (in columns, with $\lambda = 0$ corresponding to the null) at the 5% nominal size for DGP P5, at horizons $h = 1$ and $h = 5$. The sample size is $P = 2500$ in all simulations, and for $h = 1$ the asymptotic critical values reported in Table 1 are used, while for $h = 5$, the critical values are based on 5000 bootstrap simulations. The number of Monte Carlo simulations is 10,000. The domain of r is discretized over $[0 : 0.001 : 1]$.

4.3 An illustrative example

DGP P6: GARCH(1,1) versus Student-*t*. Empirically, it is often difficult to distinguish whether innovations are heavy-tailed or they feature time-varying volatility. To illustrate the detection capabilities of the tests in such a situation, as an additional example we take the GARCH(1,1) model of DGP S7 as the true model, and vary the α parameter responsible for volatility dynamics, regulating the degree of misspecification. We use a misspecified predictive distribution based on a rescaled Student-t distribution, where parameters are set to match the first four moments of the GARCH(1,1) model's unconditional distribution.[12] When $\alpha = 0$, the conditional variance σ_t^2 is a deterministic time-varying process. If in addition one sets $\sigma_0^2 = \omega/(1-\beta)$, starting at the stationary value, the GARCH(1,1) process becomes a Gaussian white noise, whose predictive CDF is identical to that of the parametrized Student-t (with $\nu \to \infty$). For a given $\beta > 0$, as α increases, the variance, kurtosis and the autocorrelation of y_t^2 increase. The matched Student-t is able to capture the increase in variance and kurtosis but not the increasingly persistent volatility dynamics and other potentially relevant features of the distribution (*e.g.*, higher order moments).

Table 8 shows the empirical rejection rates. In the absence of misspecification ($\alpha = 0$), all tests reject at the nominal size of 5% or very close to it. As the degree of misspecification increases with α, the tests reject more frequently, reaching essentially perfect rates when $\alpha = 0.12$.

Table 8: Rejection rates: DGP P6

Panel A: DGP P6, $h=1$		$\alpha=0.00$		$\alpha=0.02$		$\alpha=0.06$		$\alpha=0.08$		$\alpha=0.10$		$\alpha=0.12$	
Region or weight function		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0,1]$	0.05	0.05	0.05	0.05	0.12	0.11	0.34	0.36	0.83	0.85	1.00	1.00
Left tail	$r \in [0,0.1]$	0.05	0.05	0.07	0.07	0.20	0.22	0.49	0.52	0.87	0.88	1.00	1.00
Right tail	$r \in [0.9,1]$	0.05	0.05	0.06	0.06	0.20	0.22	0.50	0.52	0.87	0.87	1.00	1.00
Left half	$r \in [0,0.5]$	0.05	0.05	0.05	0.05	0.10	0.11	0.26	0.29	0.71	0.73	0.99	0.99
Right half	$r \in [0.5,1]$	0.05	0.05	0.05	0.06	0.09	0.10	0.26	0.29	0.71	0.73	1.00	1.00
Center	$r \in [0.1,0.9]$	0.05	0.05	0.05	0.05	0.12	0.10	0.34	0.31	0.82	0.82	1.00	1.00
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.05	0.04	0.07	0.07	0.26	0.30	0.60	0.64	0.93	0.94	1.00	1.00
Left tail	$w(r) = (1-r)^2$	0.05	0.05	0.06	0.05	0.18	0.12	0.48	0.36	0.87	0.82	1.00	1.00
Right tail	$w(r) = r^2$	0.05	0.05	0.06	0.05	0.18	0.12	0.48	0.36	0.87	0.82	1.00	1.00
Center	$w(r) = r(1-r)$	0.05	0.05	0.05	0.05	0.07	0.08	0.15	0.25	0.54	0.77	0.99	1.00
Tails	$w(r) = (2r-1)^2$	0.05	0.05	0.07	0.06	0.27	0.24	0.61	0.60	0.93	0.94	1.00	1.00
Panel B: DGP P6, $h=5$		$\alpha=0.00$		$\alpha=0.02$		$\alpha=0.06$		$\alpha=0.08$		$\alpha=0.10$		$\alpha=0.12$	
Region or weight function		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0,1]$	0.05	0.05	0.05	0.05	0.10	0.09	0.30	0.30	0.78	0.81	1.00	1.00
Left tail	$r \in [0,0.1]$	0.05	0.05	0.06	0.06	0.14	0.15	0.36	0.37	0.78	0.78	0.99	0.99
Right tail	$r \in [0.9,1]$	0.05	0.05	0.06	0.06	0.14	0.15	0.36	0.38	0.79	0.79	0.99	0.99
Left half	$r \in [0,0.5]$	0.05	0.05	0.05	0.05	0.09	0.09	0.23	0.23	0.68	0.67	0.99	0.99
Right half	$r \in [0.5,1]$	0.05	0.05	0.05	0.05	0.08	0.09	0.23	0.24	0.68	0.68	0.99	0.99
Center	$r \in [0.1,0.9]$	0.05	0.05	0.05	0.05	0.10	0.08	0.30	0.26	0.78	0.78	1.00	1.00
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.05	0.05	0.06	0.06	0.17	0.20	0.44	0.49	0.85	0.87	1.00	1.00
Left tail	$w(r) = (1-r)^2$	0.05	0.05	0.06	0.05	0.14	0.10	0.36	0.29	0.80	0.77	1.00	1.00
Right tail	$w(r) = r^2$	0.05	0.05	0.06	0.05	0.14	0.10	0.37	0.30	0.81	0.77	1.00	1.00
Center	$w(r) = r(1-r)$	0.05	0.05	0.05	0.05	0.07	0.07	0.14	0.22	0.53	0.72	0.99	1.00
Tails	$w(r) = (2r-1)^2$	0.05	0.05	0.06	0.06	0.17	0.17	0.44	0.49	0.85	0.90	1.00	1.00

Note: The table reports the empirical rejection rates of the Kolmogorov–Smirnov- (KS) and Cramér–von Mises- (CvM) type tests in various regions and with various weights (in rows) for different strengths of volatility dynamic, denoted by α (in columns, with $\alpha = 0$ corresponding to no volatility dynamics) at the 5% nominal size for DGP P6, at horizons $h = 1$ and $h = 5$. The sample size is $P = 2500$ in all simulations, and for $h = 1$ the asymptotic critical values reported in Table 1 are used, while for $h = 5$, the critical values are based on 5000 bootstrap simulations. The number of Monte Carlo simulations is 10,000. The domain of r is discretized over $[0 : 0.001 : 1]$.

[12] In our exercises, the parameters satisfy $3\alpha^2 + 2\alpha\beta + \beta^2 < 1$, hence the fourth moment of y_t exists.

5 Empirical application

Starting with the seminal paper by Stock and Watson (2007), unobserved components (UC) models have proven to provide accurate inflation forecasts (see, *e.g.*, Stock and Watson (2016) for a multivariate extension, and Chan (2013) and Zhang et al. (2020) for further specification choices). As a novelty, in our empirical application, we use UC models to forecast US industrial production growth. These models decompose the observed series y_t into a random walk permanent component τ_t and a stationary component ε_t^y as

$$y_t = \tau_t + \varepsilon_t^y \,, \tag{25}$$

$$\tau_t = \tau_{t-1} + \varepsilon_t^\tau \,. \tag{26}$$

The first model, labeled as **UC**, treats ε_t^y and ε_t^τ as *iid* normal innovations, also independent of each other contemporaneously and at all leads and lags:

$$\varepsilon_t^y \overset{iid}{\sim} \mathcal{N}(0, \sigma_y^2) \,, \tag{27}$$

$$\varepsilon_t^\tau \overset{iid}{\sim} \mathcal{N}(0, \sigma_\tau^2) \,. \tag{28}$$

The second model, labeled as **UC-SV**, replaces Equation (27) by specifying a stationary AR(1) stochastic volatility (SV) process h_t for the log variance of ε_t^y given by

$$\varepsilon_t^y = \exp\left(\frac{h_t}{2}\right) \cdot \varepsilon_t^v \,, \qquad \varepsilon_t^v \overset{iid}{\sim} \mathcal{N}(0,1) \,, \tag{29}$$

$$h_t = \mu_h + \phi_h (h_{t-1} - \mu_h) + \varepsilon_t^h \,, \qquad \varepsilon_t^h \overset{iid}{\sim} \mathcal{N}(0, \sigma_h^2) \,, \tag{30}$$

where ε_t^v, ε_t^h and ε_t^τ are independent of each other contemporaneously and at all leads and lags.

To illustrate the performance of our proposed method, we estimated the models in a recursive fashion, using both monthly and quarterly seasonally adjusted US industrial production from the 2025 March vintages of the FRED-MD (McCracken and Ng, 2016) and FRED-QD (McCracken and Ng, 2021) databases, respectively.[13] In particular, we used the annualized monthly or quarterly logarithmic growth rates of the industrial production indices (mnemonics: `INDPRO`), given by

$$y_t = f \cdot 100 \cdot \left[\log\left(\text{INDPRO}_t\right) - \log\left(\text{INDPRO}_{t-1}\right)\right] \,, \tag{31}$$

[13]The FRED-MD and FRED-QD databases are available on the the St. Louis Fed's website at https://www.stlouisfed.org/research/economists/mccracken/fred-databases.

where $f = 12$ for monthly data, and $f = 4$ for quarterly data.

In the case of monthly frequency, the first estimation window spans from 1959:M2 to 1974:M12, and we predict industrial production growth $h = \{1, 3, 6\}$ months and $h = 1$ year ahead, corresponding to month-on-month (MoM) and year-on-year (YoY) growth rates, with target dates of 1975:M1, 1975:M3, 1975:M6 and 1975:M12, respectively. Next, we recursively extend the estimation sample by one month, to 1959:M2 – 1975:M1, and again generate forecasts $h = \{1, 3, 6\}$ months and $h = 1$ year ahead. We proceed in this recursive manner until reaching 2024:M11. The evaluation sample hence spans from 1974:M12 $+ h$ to 2024:M12, resulting in $P = 601 - h$ out-of-sample observations. Similarly, when estimating the models at the quarterly frequency, the first estimation window covers the period between 1959:Q2 and 1974:Q4, and we forecast $h = \{1, 2, 4\}$ quarters and $h = 1$ year ahead, corresponding to quarter-on-quarter (QoQ) and year-on-year growth rates, respectively. Next, the estimation window is recursively extended by one quarter at a time until reaching 2024:Q3. The corresponding evaluation period hence spans from 1974:Q4 $+ h$ to 2024:Q4, with $P = 201 - h$ out-of-sample observations.

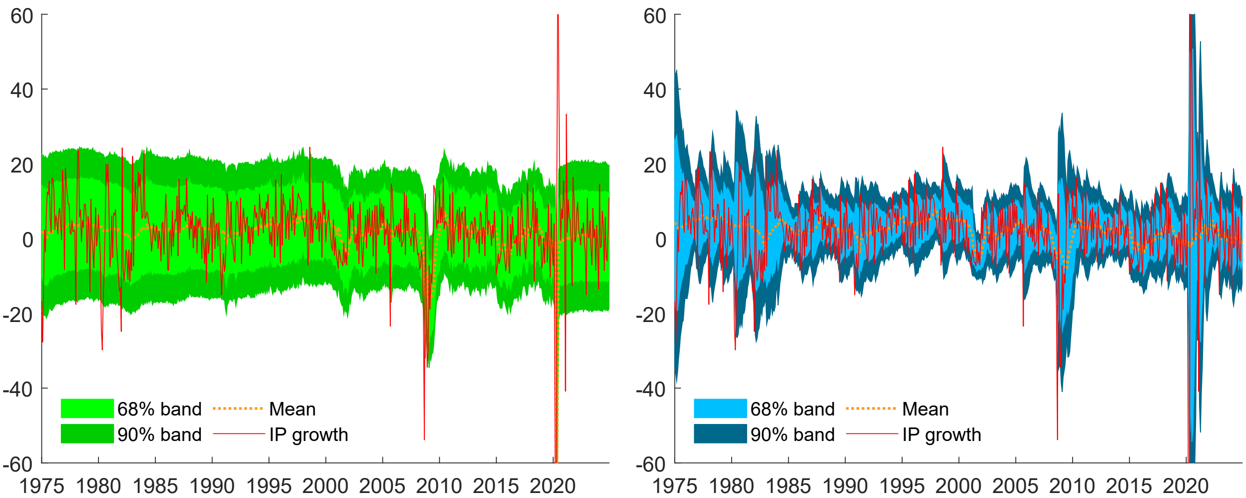
We estimated both models using Markov-Chain Monte Carlo methods building on Chan and Jeliazkov (2009) and Chan and Hsiao (2014). For details on the prior specifications and the posterior samplers, see Appendix D. Additional empirical results are collected in Appendix E.

Figure 2 shows the equal-tailed 68% and 90% predictive bands corresponding to the one-month-ahead forecasts of the **UC** and **UCSV** models. It is immediately clear that the **UC** model's predictive bands in Panel (a) are of largely constant width, while the **UCSV** model's stochastic volatility component generates sizable time-variation, as Panel (b) demonstrates. Turning to the quarterly frequency predictive bands, we can see the same pattern in Figure 3.

Table 9 shows the formal evaluation of the correct calibration of the predictive densities of the **UC** and **UCSV** models estimated on monthly frequency data. As we can see, the **UC** model does not produce correctly calibrated density forecasts at any of the monthly horizons (columns "1M", "3M" and "6M") in any part of the predictive distribution (in rows), while in the case of the **UCSV** model we cannot reject the null hypothesis of correct calibration at the 5% significance level. Interestingly, however, at the annual horizon (forecasting year-on-year growth), the **UC** model delivers correctly calibrated density forecasts. This can be explained by the fact that year-on-year growth rates evolve in a much smoother manner, which mitigates the need for stochastic volatility.

Similarly, in the case of quarterly data, in Table 10 we see that the **UCSV** model produces correctly calibrated density forecasts at all horizons and in all regions of the predictive distribution.

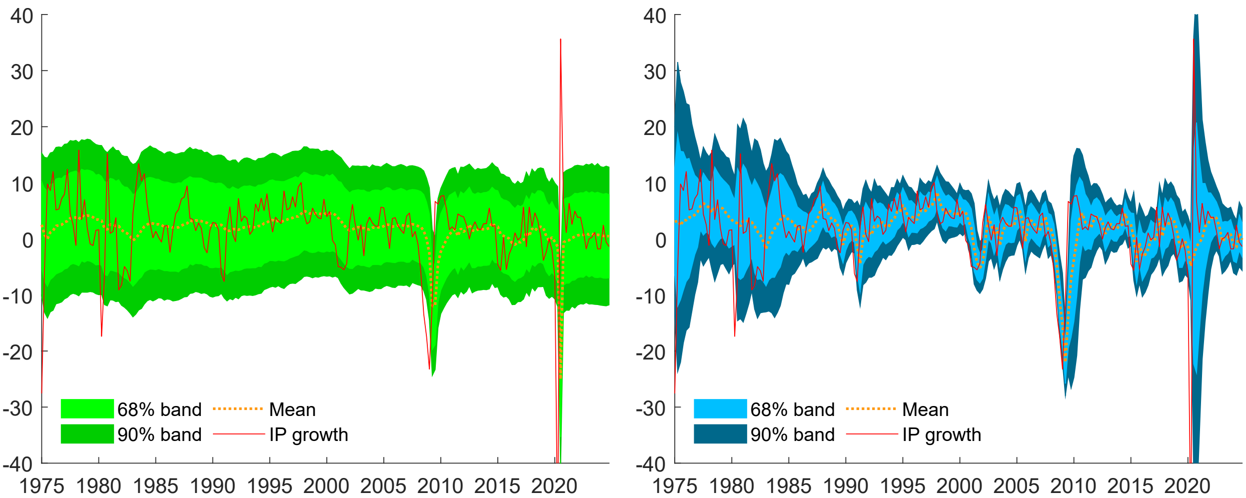
Figure 2: One-month-ahead equal-tailed 68% and 90% predictive bands



(a) **UC** model (b) **UCSV** model

Note: The panels show the 68% and 90% predictive bands (with light and dark shades, respectively) along with the mean predictions (dotted lines) and the realized annualized month-on-month industrial production growth (solid lines) between 1975:M1 and 2024:M12. Dates on the horizontal axes correspond to the forecast target dates.

Figure 3: One-quarter-ahead equal-tailed 68% and 90% predictive bands



(a) **UC** model (b) **UCSV** model

Note: The panels show the 68% and 90% predictive bands (with light and dark shades, respectively) along with the mean predictions (dotted lines) and the realized annualized quarter-on-quarter industrial production growth (solid lines) between 1975:Q1 and 2024:Q4. Dates on the horizontal axes correspond to the forecast target dates.

The **UC** model's quarterly forecasts are mostly incorrectly calibrated, although its one-year-ahead year-on-year forecasts (columns labeled "1Y") do not suffer from miscalibration.

Figure 4 illustrates our findings for the case of one-month-ahead forecasts by comparing the empirical CDF of the PITs to the 45 degree line. Following Rossi and Sekhposyan (2019), to help visual interpretation, we also display the critical value bands defined as $r \pm \widehat{c}_{\kappa,P}/\sqrt{P}$, where $\widehat{c}_{\kappa,P}$ is the 5% critical value of the Kolmogorov–Smirnov-test. If the empirical CDF crosses the bands, the null hypothesis of correct calibration over $r \in [0,1]$ is rejected. Considering the **UC** model in

Table 9: Correct specification tests for monthly US industrial production growth

		UC								**UCSV**							
		KS				CvM				KS				CvM			
Region or weight function		1M	3M	6M	1Y	1M	3M	6M	1Y	1M	3M	6M	1Y	1M	3M	6M	1Y
Full	$r \in [0,1]$	0.00	0.00	0.00	**0.24**	0.00	0.00	0.00	**0.46**	**0.13**	**0.48**	**0.65**	**0.14**	**0.18**	**0.42**	**0.59**	**0.28**
Left tail	$r \in [0,0.1]$	0.00	0.00	0.00	**0.57**	0.00	0.00	0.00	**0.62**	**0.06**	**0.25**	**0.22**	0.03	**0.06**	**0.27**	**0.18**	0.02
Right tail	$r \in [0.9,1]$	0.00	0.00	0.00	**0.21**	0.00	0.00	0.01	**0.36**	**0.78**	**0.54**	**0.85**	**0.32**	**0.64**	**0.64**	**0.89**	**0.25**
Left half	$r \in [0,0.5]$	0.00	0.00	0.00	**0.17**	0.00	0.00	0.00	**0.24**	**0.08**	**0.38**	**0.52**	**0.12**	**0.08**	**0.28**	**0.42**	**0.16**
Right half	$r \in [0.5,1]$	0.00	0.00	0.01	**0.78**	0.00	0.00	0.00	**0.85**	**0.65**	**0.76**	**0.62**	**0.74**	**0.50**	**0.71**	**0.78**	**0.76**
Center	$r \in [0.1,0.9]$	0.00	0.00	0.00	**0.24**	0.00	0.00	0.00	**0.45**	**0.13**	**0.48**	**0.64**	**0.15**	**0.18**	**0.42**	**0.60**	**0.37**
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.00	0.00	0.00	**0.37**	0.00	0.00	0.00	**0.57**	**0.13**	**0.35**	**0.26**	0.03	**0.16**	**0.40**	**0.30**	0.02
Left tail	$w(r) = (1-r)^2$	0.00	0.00	0.00	**0.29**	0.00	0.00	0.00	**0.31**	**0.08**	**0.33**	**0.29**	0.02	**0.09**	**0.31**	**0.41**	**0.13**
Right tail	$w(r) = r^2$	0.00	0.00	0.00	**0.29**	0.00	0.00	0.00	**0.66**	**0.61**	**0.78**	**0.79**	**0.47**	**0.41**	**0.62**	**0.80**	**0.63**
Center	$w(r) = r(1-r)$	0.00	0.01	0.01	**0.23**	0.00	0.00	0.00	**0.44**	**0.23**	**0.47**	**0.70**	**0.49**	**0.19**	**0.42**	**0.62**	**0.40**
Tails	$w(r) = (2r-1)^2$	0.00	0.00	0.00	**0.20**	0.00	0.00	0.00	**0.50**	**0.16**	**0.42**	**0.22**	0.02	**0.14**	**0.41**	**0.43**	**0.09**

Note: The table reports the p-values of the correct specification of predictive densities of the UC and UCSV models based on the Kolmogorov–Smirnov- and Cramér–von Mises-type (KS and CvM, respectively) test statistics in various regions or with various weight functions (in rows) at forecast horizons of $h = 1,3,6$ months (annualized MoM growth) and $h = 1$ year (YoY growth) in columns. Boldface indicates entries for which the null hypothesis of correct calibration is not rejected at the $\alpha = 5\%$ significance level. The forecast evaluation period spans from 1974:M12 + h to 2024:M12. For $h = 1$, the p-values are computed using the limiting Gaussian process with 10,000 simulations. For $h > 1$, the critical values are based on the bootstrap with block length $\ell = \max\{h-1, \lfloor P^{1/3} \rfloor\}$. The number of bootstrap replications is 10,000, the domain of r is discretized over [0 : 0.001 : 1].

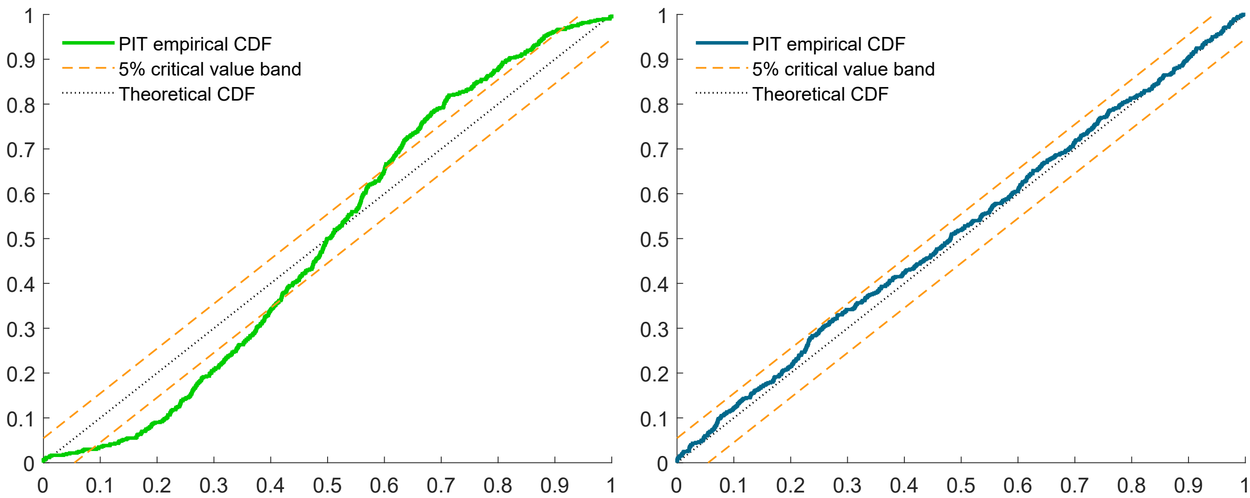
Table 10: Correct specification tests for quarterly US industrial production growth

		UC								**UCSV**							
		KS				CvM				KS				CvM			
Region or weight function		1Q	2Q	4Q	1Y	1Q	2Q	4Q	1Y	1Q	2Q	4Q	1Y	1Q	2Q	4Q	1Y
Full	$r \in [0,1]$	0.00	0.01	0.01	**0.44**	0.00	0.00	0.01	**0.53**	**0.58**	**0.83**	**0.76**	**0.43**	**0.58**	**0.80**	**0.93**	**0.60**
Left tail	$r \in [0,0.1]$	0.01	0.01	**0.12**	**0.47**	0.02	0.03	**0.15**	**0.54**	**0.43**	**0.25**	**0.30**	**0.12**	**0.41**	**0.16**	**0.34**	**0.09**
Right tail	$r \in [0.9,1]$	0.01	0.02	0.03	**0.34**	0.01	**0.09**	0.02	**0.41**	**0.39**	**0.66**	**0.92**	**0.36**	**0.37**	**0.52**	**0.93**	**0.36**
Left half	$r \in [0,0.5]$	0.00	0.00	0.00	**0.31**	0.00	0.01	0.01	**0.32**	**0.66**	**0.70**	**0.83**	**0.34**	**0.61**	**0.71**	**0.86**	**0.38**
Right half	$r \in [0.5,1]$	0.00	0.05	**0.07**	**0.87**	0.00	0.02	0.04	**0.79**	**0.42**	**0.65**	**0.58**	**0.90**	**0.45**	**0.74**	**0.87**	**0.93**
Center	$r \in [0.1,0.9]$	0.00	0.01	0.01	**0.44**	0.00	0.00	0.01	**0.52**	**0.58**	**0.81**	**0.75**	**0.42**	**0.58**	**0.89**	**0.94**	**0.70**
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.02	0.02	**0.10**	**0.64**	0.00	0.03	0.04	**0.60**	**0.62**	**0.31**	**0.47**	**0.13**	**0.47**	**0.19**	**0.56**	**0.12**
Left tail	$w(r) = (1-r)^2$	0.00	0.01	0.01	**0.44**	0.00	0.01	0.01	**0.42**	**0.42**	**0.30**	**0.42**	**0.14**	**0.61**	**0.65**	**0.85**	**0.34**
Right tail	$w(r) = r^2$	0.00	0.00	0.01	**0.64**	0.00	0.01	0.01	**0.65**	**0.68**	**0.69**	**0.56**	**0.56**	**0.53**	**0.81**	**0.92**	**0.90**
Center	$w(r) = r(1-r)$	0.00	0.01	0.01	**0.39**	0.00	0.01	0.01	**0.50**	**0.45**	**0.82**	**0.82**	**0.86**	**0.55**	**0.85**	**0.95**	**0.73**
Tails	$w(r) = (2r-1)^2$	0.01	0.04	0.04	**0.58**	0.00	0.00	0.00	**0.64**	**0.57**	**0.23**	**0.33**	**0.13**	**0.66**	**0.60**	**0.82**	**0.30**

Note: The table reports the p-values of the correct specification of predictive densities of the UC and UCSV models based on the Kolmogorov–Smirnov- and Cramér–von Mises-type (KS and CvM, respectively) test statistics in various regions or with various weight functions (in rows) at forecast horizons of $h = 1,2,4$ quarters (annualized QoQ growth) and $h = 1$ year (YoY growth) in columns. Boldface indicates entries for which the null hypothesis of correct calibration is not rejected at the $\alpha = 5\%$ significance level. The forecast evaluation period spans from 1974:Q4 + h to 2024:Q4. For $h = 1$, the p-values are computed using the finite-sample standard uniform distribution with 10,000 simulations. For $h > 1$, the critical values are based on the bootstrap with block length $\ell = \max\{h-1, \lfloor P^{1/3} \rfloor\}$. The number of bootstrap replications is 10,000, the domain of r is discretized over [0 : 0.001 : 1].

Panel (a), the empirical cumulative distribution function of the PITs crosses the 5% critical value band based on the Kolmogorov–Smirnov-type test, indicating the rejection of the null hypothesis. On the other hand, the empirical CDF of the **UCSV** model's PITs is inside the band in Panel (b), close to the 45 degree line, hence we fail to reject the null of correct calibration. Figure 5 shows a similar pattern for quarterly data.

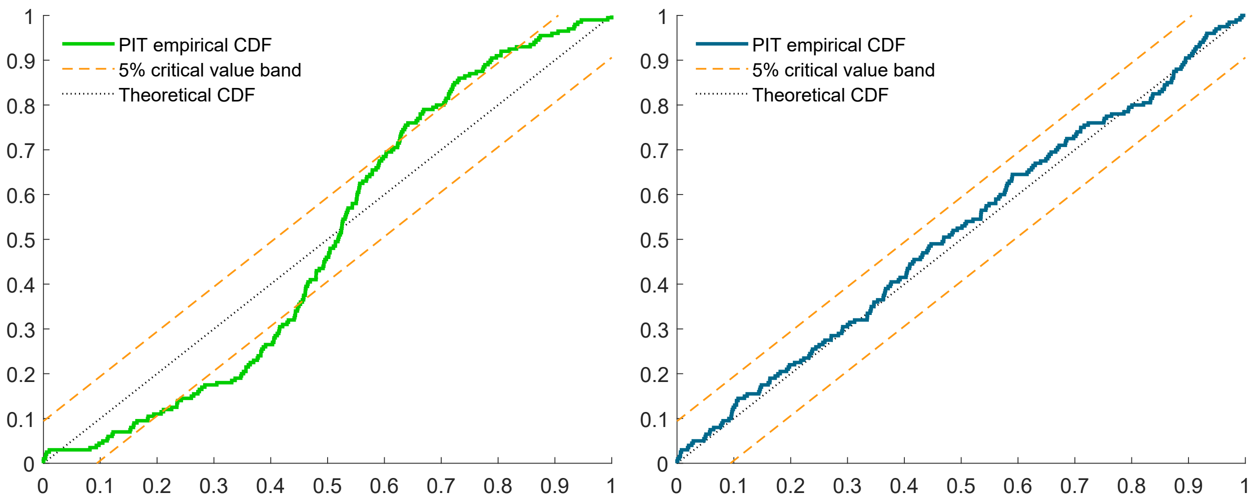
Figure 4: Empirical CDFs and 5% critical value bands of one-month-ahead PITs



(a) **UC** model (b) **UCSV** model

Note: The figures show the empirical CDFs of the PITs (solid lines), the 5% critical bands based on the Kolmogorov–Smirnov-type test (dashed lines), and the 45 degree lines (dotted) corresponding to the theoretical CDF under correct calibration.

Figure 5: Empirical CDFs and 5% critical value bands of one-quarter-ahead PITs



(a) **UC** model (b) **UCSV** model

Note: The figures show the empirical CDFs of the PITs (solid lines), the 5% critical bands based on the Kolmogorov–Smirnov-type test (dashed lines), and the 45 degree lines (dotted) corresponding to the theoretical CDF under correct calibration.

To summarize, our empirical results contribute to and are in line with the large literature highlighting sizable forecasting gains from stochastic volatility when predicting US macroeconomic aggregates (see, *e.g.*, Clark (2011), Clark and Ravazzolo (2015) and Chiu et al. (2017)).

6 Conclusion

This paper introduces a simplified yet powerful framework for testing the correct specification of conditional predictive densities using the probability integral transform. By treating the PIT as a primitive we can relax restrictive assumptions found in the earlier literature. In particular, we extend the tests proposed by Rossi and Sekhposyan (2019) to a broader class of models, including those not satisfying strong mixing conditions and whose predictive conditional distribution can depend upon the entire history of the process, such as those featuring stochastic volatility or other latent dynamic structures.

Monte Carlo simulations confirm that the proposed tests exhibit desirable size and power properties across a wide range of data-generating processes, including non-linear and non-Gaussian settings. Weighted Kolmogorov–Smirnov- and Cramér–von Mises-type statistics targeting the entire distribution as well as parts of it are considered. An empirical application to US industrial production growth illustrates the practical value of the test. We show that incorporating stochastic volatility into an unobserved components model is essential for generating correctly calibrated density forecasts of US industrial production growth at both monthly and quarterly frequencies.

Overall, this work provides both theoretical and empirical justification for PIT-based tests in modern forecasting environments, offering a flexible tool for researchers and practitioners to evaluate the reliability of density forecasts in complex real-world settings. Potential further applications include the evaluation of fan charts of GDP growth and inflation, and various “at-risk” measures, such as growth-at-risk or house prices-at-risk.

References

Adrian, Tobias, Nina Boyarchenko and Domenico Giannone. (2019). "Vulnerable growth". *American Economic Review*, 109(4), pp. 1263-1289. https://doi.org/10.1257/aer.20161923

Adrian, Tobias, Nina Boyarchenko and Domenico Giannone. (2021). "Multimodality in macrofinancial dynamics". *International Economic Review*, 62(2), pp. 861-886. https://doi.org/10.1111/iere.12501

Amisano, Gianni, and Raffaella Giacomini. (2007). "Comparing density forecasts via weighted likelihood ratio tests". *Journal of Business & Economic Statistics*, 25(2), pp. 177-190. https://doi.org/10.1198/073500106000000332

Andersen, Torben G., Luca Benzoni and Jesper Lund. (2002). "An empirical investigation of continuous-time equity return models". *The Journal of Finance*, 57(3), pp. 1239-1284. https://doi.org/10.1111/1540-6261.00460

Bai, Jushan. (2003). "Testing parametric conditional distributions of dynamic models". *The Review of Economics and Statistics*, 85(3), pp. 531-549. https://doi.org/10.1162/003465303322369704

Bai, Jushan, and Zhihong Chen. (2008). "Testing multivariate distributions in GARCH models". *Journal of Econometrics*, 143(1), pp. 19-36. https://doi.org/10.1016/j.jeconom.2007.08.012

Billingsley, Patrick. (1968). *Convergence of Probability Measures*. Wiley. https://cermics.enpc.fr/~monneau/Billingsley-2eme-edition.pdf

Bollerslev, Tim. (1986). "Generalized autoregressive conditional heteroskedasticity". *Journal of Econometrics*, 31(3), pp. 307-327. https://doi.org/10.1016/0304-4076(86)90063-1

Brownlees, Christian, and André B. M. Souza. (2021). "Backtesting global Growth-at-Risk". *Journal of Monetary Economics*, 118, pp. 312-330. https://doi.org/10.1016/j.jmoneco.2020.11.003

Casella, George, and Roger L. Berger. (2002). *Statistical Inference*. 2nd ed. Thomson Learning. https://doi.org/10.1201/9781003456285

Chan, Joshua C. C. (2013). "Moving average stochastic volatility models with application to inflation forecast". *Journal of Econometrics*, 176(2), pp. 162-172. https://doi.org/10.1016/j.jeconom.2013.05.003

Chan, Joshua C. C., and Cody Y. L. Hsiao. (2014). "Chapter 6. Estimation of stochastic volatility models with heavy tails and serial dependence". In Jeliazkov, Ivan, and Xin-She Yang (eds.), *Bayesian Inference in the Social Sciences*, pp. 155-176. https://doi.org/10.1002/9781118771051.ch6

Chan, Joshua C. C., and Ivan Jeliazkov. (2009). "Efficient simulation and integrated likelihood estimation in state space models". *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2), pp. 101-120. https://www.inderscienceonline.com/doi/abs/10.1504/IJMMNO.2009.03009

Chiu, Ching-Wai (Jeremy), Haroon Mumtaz and Gábor Pintér. (2017). "Forecasting with VAR models: Fat tails and stochastic volatility". *International Journal of Forecasting*, 33(4), pp. 1124-1143. https://doi.org/10.1016/j.ijforecast.2017.03.001

Clark, Todd E. (2011). “Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility”. *Journal of Business & Economic Statistics,* 29(3), pp. 327-341. https://doi.org/10.1198/jbes.2010.09248

Clark, Todd E., Florian Huber, Gary Koop and Massimiliano Marcellino. (2024). “Forecasting US inflation using Bayesian nonparametric models”. *The Annals of Applied Statistics,* 18(2), pp. 1421-1444. https://doi.org/10.1214/23-AOAS1841

Clark, Todd E., and Francesco Ravazzolo. (2015). “Macroeconomic forecasting performance under alternative specifications of time-varying volatility”. *Journal of Applied Econometrics,* 30(4), pp. 551-575. https://doi.org/10.1002/jae.2379

Corradi, Valentina, and Norman R. Swanson. (2006a). “Bootstrap conditional distribution tests in the presence of dynamic misspecification”. *Journal of Econometrics,* 133(2), pp. 779-806. https://doi.org/10.1016/j.jeconom.2005.06.013

Corradi, Valentina, and Norman R. Swanson. (2006b). “Predictive density and conditional confidence interval accuracy tests”. *Journal of Econometrics,* 135(1-2), pp. 187-228. https://doi.org/10.1016/j.jeconom.2005.07.026

Corradi, Valentina, and Norman R. Swanson. (2006c). “Chapter 5 Predictive density evaluation”. In Elliott, Graham, Clive W. J. Granger and Allan Timmermann (eds.), *Handbook of Economic Forecasting,* 1. Elsevier, pp. 197-284. https://doi.org/10.1016/S1574-0706(05)01005-0

Demetrescu, Matei, Felix Kießner and Malte Knüppel. (2025). *Simultaneous inference bands for the PIT histogram.* International Symposium of Forecasting 2024, Dijon, France, June 30-July 3. https://isf.forecasters.org/wp-content/uploads/BookOfAbstracts-ISF2024_FINAL.pdf

Diebold, Francis X. (2015). “Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests”. *Journal of Business & Economic Statistics,* 33(1), pp. 1-9. https://doi.org/10.1080/07350015.2014.983236

Diebold, Francis X., Todd A. Gunther and Anthony S. Tay. (1998). “Evaluating density forecasts with applications to financial risk management”. *International Economic Review,* 39(4), pp. 863-883. https://doi.org/10.2307/2527342

Diebold, Francis X., Jinyong Hahn and Anthony S. Tay. (1999). “Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns on foreign exchange”. *The Review of Economics and Statistics,* 81(4), pp. 661-673. https://doi.org/10.1162/003465399558526

Dovern, Jonas, and Hans Manner. (2020). “Order-invariant tests for proper calibration of multivariate density forecasts”. *Journal of Applied Econometrics,* 35(4), pp. 440-456. https://doi.org/10.1002/jae.2755

Engle, Robert F. (1982). “Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation”. *Econometrica,* 50(4), pp. 987-1007. https://doi.org/10.2307/1912773

Gneiting, Tilmann, and Matthias Katzfuss. (2014). “Probabilistic forecasting”. *Annual Review of Statistics and Its Application,* 1(1), pp. 125-151. https://doi.org/10.1146/annurev-statistics-062713-085831

Gneiting, Tilmann, and Roopesh Ranjan. (2011). “Comparing density forecasts using threshold- and quantile-weighted scoring rules”. *Journal of Business & Economic Statistics*, 29(3), pp. 411-422. https://doi.org/10.1198/jbes.2010.08110

Hansen, Peter R., and Asger Lunde. (2005). “A forecast comparison of volatility models: does anything beat a GARCH(1,1)?”. *Journal of Applied Econometrics*, 20(7), pp. 873-889. https://doi.org/10.1002/jae.800

Inoue, Atsushi. (2001). “Testing for distributional change in time series”. *Econometric Theory*, 17(1), pp. 156-187. https://doi.org/10.1017/S0266466601171057

International Monetary Fund. (2019). *Global Financial Stability Report, April 2019: Vulnerabilities in a Maturing Credit Cycle*. https://www.imf.org/en/Publications/GFSR/Issues/2019/03/27/Global-Financial-Stability-Report-April-2019?utm_source=burnaby%20now&utm_campaign=burnaby%20now&utm_medium=referral

Janson, Svante. (2021). “A central limit theorem for *m*-dependent variables”. *arXiv*, math.PR(2108.12263). https://doi.org/10.48550/arXiv.2108.12263

Knüppel, Malte. (2015). “Evaluating the calibration of multi-step-ahead density forecasts using raw moments”. *Journal of Business & Economic Statistics*, 33(2), pp. 270-281. https://doi.org/10.1080/07350015.2014.948175

Knüppel, Malte, Fabian Krüger and Marc-Oliver Pohle. (2023). “Score-based calibration testing for multivariate forecast distributions”. *arXiv*, econ.EM(2211.16362). https://doi.org/10.48550/arXiv.2211.16362

Lhuissier, Stéphane. (2022). “Financial conditions and macroeconomic downside risks in the euro area”. *European Economic Review*, 143(104046). https://doi.org/10.1016/j.euroecorev.2022.104046

McCracken, Michael W., and Serena Ng. (2016). “FRED-MD: A monthly database for macroeconomic research”. *Journal of Business & Economic Statistics*, 34(4), pp. 574-589. https://doi.org/10.1080/07350015.2015.1086655

McCracken, Michael W., and Serena Ng. (2021). “FRED-QD: A quarterly database for macroeconomic research”. *Federal Reserve Bank of St. Louis Review*, 103(1), pp. 1-44. https://doi.org/10.20955/r.103.1-44

Merton, Robert C. (1976). “Option pricing when underlying stock returns are discontinuous”. *Journal of Financial Economics*, 3(1-2), pp. 125-144. https://doi.org/10.1016/0304-405X(76)90022-2

Omori, Yasuhiro, Siddhartha Chib, Neil Shephard and Jouchi Nakajima. (2007). “Stochastic volatility with leverage: Fast and efficient likelihood inference”. *Journal of Econometrics*, 140(2), pp. 425-449. https://doi.org/10.1016/j.jeconom.2006.07.008

Rossi, Barbara, and Tatevik Sekhposyan. (2019). “Alternative tests for correct specification of conditional predictive densities”. *Journal of Econometrics*, 208(2), pp. 638-657. https://doi.org/10.1016/j.jeconom.2018.07.008

Sklar, Abe. (1959). “Fonctions de répartition à N dimensions et leurs marges”. *Publications de l’Institut de Statistique de l’Université de Paris*, 8, pp. 229-231. https://hal.science/hal-04094463v1

Stock, James H., and Mark W. Watson. (2007). "Why has U.S. inflation become harder to forecast?". *Journal of Money, Credit and Banking*, 39(s1), pp. 3-33. https://doi.org/10.1111/j.1538-4616.2007.00014.x

Stock, James H., and Mark W. Watson. (2016). "Core inflation and trend inflation". *The Review of Economics and Statistics*, 98(4), pp. 770-784. https://doi.org/10.1162/REST_a_00608

Timmermann, Allan. (2000). "Editorial: Density forecasting in economics and finance". *Journal of Forecasting*, 19(4), pp. 231-234. https://doi.org/10.1002/1099-131X(200007)19:4%3C231::AID-FOR771%3E3.0.CO;2-%23

Timmermann, Allan, and Graham Elliott. (2016). *Economic Forecasting*. Princeton University Press.

Zhang, Bo, Joshua C. C. Chan and Jamie L. Cross. (2020). "Stochastic volatility models with ARMA innovations: An application to G7 inflation forecasts". *International Journal of Forecasting*, 36(4), pp. 1318-1328. https://doi.org/10.1016/j.ijforecast.2020.01.004

Appendices

Appendix A Proofs

This appendix presents the proofs of lemmas and theorems reported in the paper.

Proof of Lemma 1. Define the quantile function as $\Phi_0^{-1}(r \mid \mathscr{F}_t) \equiv \inf\{y : \Phi_0(y \mid \mathscr{F}_t) \geq r\}$ for $r \in (0,1)$, $\Phi_0^{-1}(1 \mid \mathscr{F}_t) = \infty$ and $\Phi_0^{-1}(0 \mid \mathscr{F}_t) = -\infty$ (see Casella and Berger, 2002, Theorem 2.1.10, p. 54). Consider

$$\mathbb{P}(\Phi_0(y_{t+h} \mid \mathscr{F}_t) \leq y) = \mathbb{P}(\Phi_0^{-1}(\Phi_0(y_{t+h} \mid \mathscr{F}_t) \mid \mathscr{F}_t) \leq \Phi_0^{-1}(y \mid \mathscr{F}_t)) \tag{A.1}$$

$$= \mathbb{P}(y_{t+h} \leq \Phi_0^{-1}(y \mid \mathscr{F}_t)) \tag{A.2}$$

$$= \Phi_0(\Phi_0^{-1}(y \mid \mathscr{F}_t) \mid \mathscr{F}_t) = y\,. \tag{A.3}$$

Therefore, ξ_{t+h} follows a standard uniform distribution. Note that by using the quantile function, $\Phi_0(y_{t+h} \mid \mathscr{F}_t)$ need *not* be differentiable or invertible, only continuous. □

Proof of Lemma 2. Because ξ_{k+h} conditional on $\mathscr{F}_k$ does not depend on $\mathscr{F}_k$ (see Lemma 1), it is independent of $\mathscr{F}_k$. It follows that ξ_{k+h} is independent of ξ_{i+h} for $k \geq i+h$ since ξ_{i+h} is $\mathscr{F}_k$–measurable. The case of $i \geq k+h$ is analogous. □

Proof of Theorem 1. The proof takes inspiration from Theorem 2.1 in Inoue (2001). Weak convergence is in the space of $\mathscr{D}[0,1]$, which is the space of *càdlàg* (right-continuous with left limits) functions on $[0,1]$, in the Skorohod topology.

Step 1. The first step of the proof consists of showing that the sample covariance kernel absolutely converges to the specified covariance kernel:

$$\lim_{P\to\infty} \mathbb{E}\left[\Psi_P(r_1)\Psi_P(r_2)\right] = \sigma(r_1,r_2) \equiv \sum_{d=1-h}^{h-1} \left(F_d(r_1,r_2) - r_1 r_2\right)\,. \tag{A.4}$$

Consider

$$\mathbb{E}\left[\Psi_P(r_1)\Psi_P(r_2)\right] = \mathbb{E}\left[\left(P^{-1/2}\sum_{i=E}^{T}\left(\mathbb{1}\{\xi_{i+h} \leq r_1\} - r_1\right)\right)\left(P^{-1/2}\sum_{j=E}^{T}\left(\mathbb{1}\{\xi_{j+h} \leq r_2\} - r_2\right)\right)\right] \tag{A.5}$$

$$= \frac{1}{P}\left[\sum_{i=E}^{T}\sum_{j=E}^{T}\left(F_{j-i}(r_1,r_2) - r_1 r_2\right)\right]\,, \tag{A.6}$$

where we used the fact that, by Lemma 1, the PIT is uniformly distributed, hence $\mathbb{E}(\mathbb{1}\{\xi_{i+h} \leq r\}) = \mathbb{P}(\xi_{i+h} \leq r) = r$, and by Assumption 2, we have that $\mathbb{E}\left[\mathbb{1}\{\xi_{i+h} \leq r_1\}\mathbb{1}\{\xi_{j+h} \leq r_2\}\right] = \mathbb{E}\left[\mathbb{1}\{\xi_{i+h} \leq r_1, \xi_{j+h} \leq r_2\}\right] = \mathbb{P}\left(\xi_{i+h} \leq r_1, \xi_{j+h} \leq r_2\right) = F_{j-i}(r_1, r_2)$. Next, notice that per Lemma 2, we have $F_{|j-i|}(r_1, r_2) = r_1 r_2$ for $|j-i| \geq h$, hence

$$\frac{1}{P}\left[\sum_{i=E}^{T}\sum_{j=E}^{T}\left(F_{j-i}(r_1,r_2) - r_1 r_2\right)\right] = \sum_{d=1-h}^{h-1}\left(F_d(r_1,r_2) - r_1 r_2\right) - \frac{1}{P}\sum_{d=1-h}^{h-1}|d|\left(F_d(r_1,r_2) - r_1 r_2\right) \tag{A.7}$$

$$= \sigma(r_1, r_2) + \mathcal{O}(P^{-1}). \tag{A.8}$$

The second term in Equation (A.7) is $\mathcal{O}(P^{-1})$ since the sum is bounded by $\pm\sum_{d=1-h}^{h-1}|d|$, and thus converges to 0.

Furthermore, the left-hand side of Equation (A.4) is absolutely convergent:

$$\frac{1}{P}\sum_{i=E}^{T}\sum_{j=E}^{T}\left|\mathbb{E}\left[\left(\mathbb{1}\left\{\xi_{i+h} \leq r_1\right\} - r_1\right)\left(\mathbb{1}\left\{\xi_{j+h} \leq r_2\right\} - r_2\right)\right]\right| \tag{A.9}$$

$$\begin{aligned} &= \frac{1}{P}\sum_{i=E}^{T}\left|\mathbb{E}\left[\left(\mathbb{1}\left\{\xi_{i+h} \leq r_1\right\} - r_1\right)\left(\mathbb{1}\left\{\xi_{i+h} \leq r_2\right\} - r_2\right)\right]\right| \\ &\quad + \frac{1}{P}\sum_{i=E}^{T}\sum_{\substack{|i-j|<h \\ i \neq j}}\left|\mathbb{E}\left[\left(\mathbb{1}\left\{\xi_{i+h} \leq r_1\right\} - r_1\right)\left(\mathbb{1}\left\{\xi_{j+h} \leq r_2\right\} - r_2\right)\right]\right| \end{aligned} \tag{A.10}$$

$$\leq 1 + 2(h-1) < \infty. \tag{A.11}$$

Step 2. The next step is to show convergence of the finite-dimensional distributions of $\{\Psi_P\}$ to those of Ψ by means of the Cramér–Wold device. We need to show that

$$\sum_{j=1}^{k}\omega_j \Psi_P(r_j) \xrightarrow{d} \mathcal{N}\left(0, \lim_{P\to\infty}\mathbb{V}\mathrm{ar}\left(\sum_{j=1}^{k}\omega_j \Psi_P(r_j)\right)\right) \tag{A.12}$$

for $\forall(\omega_1, \ldots, \omega_k)' \in \mathbb{R}^k, \forall(r_1, \ldots, r_k) \in [0,1]^k$, and $\forall k \geq 1$. As the degenerate case is trivial (since it is the limit of a sequence of mean-zero normal distributions whose variance tends to zero, *i.e.*, a Dirac delta at 0), we assume $\lim_{P\to\infty}\mathbb{V}\mathrm{ar}\left(\sum_{j=1}^{k}\omega_j\Psi_P(r_j)\right) > 0$ if the limit exists.

Define the following objects:

$$X_{t+h} \equiv \frac{1}{\sqrt{P}} \sum_{j=1}^{k} \omega_j \left(\mathbb{1} \left\{ \xi_{t+h} \leq r_j \right\} - r_j \right) , \tag{A.13}$$

$$S_P \equiv \sum_{t=E}^{T} X_{t+h} = \sum_{j=1}^{k} \omega_j \Psi_P(r_j) , \tag{A.14}$$

$$\sigma^2_{S_P} \equiv \mathbb{V}\mathrm{ar}\left(S_P \right) . \tag{A.15}$$

It is sufficient to show that $\frac{S_p}{\sigma_{S_P}} \xrightarrow{d} \mathcal{N}(0,1)$. To this end we will invoke Theorem 1.1 of Janson (2021).

First we need to show that $\mathbb{E}\left(X_{t+h}\right) = 0$. We have

$$\begin{aligned}
\mathbb{E}\left(X_{t+h}\right) &= \frac{1}{\sqrt{P}} \sum_{j=1}^{k} \omega_j \, \mathbb{E}\left(\mathbb{1} \left\{ \xi_{t+h} \leq r_j \right\} - r_j \right) && \text{(A.16)}\\
&= \frac{1}{\sqrt{P}} \sum_{j=1}^{k} \omega_j \left[\mathbb{E}\left(\mathbb{1} \left\{ \xi_{t+h} \leq r_j \right\} \right) - r_j \right] && \text{(A.17)}\\
&= \frac{1}{\sqrt{P}} \sum_{j=1}^{k} \omega_j \left[r_j - r_j \right] = 0 , && \text{(A.18)}
\end{aligned}$$

where we exploited the uniformity of the PIT per Lemma 1.

Second, we need to show that $\mathbb{E}\left(X^2_{t+h}\right) < \infty$. We have

$$\begin{aligned}
X^2_{t+h} &= \frac{1}{P} \left(\sum_{j=1}^{k} \omega_j \left(\mathbb{1} \left\{ \xi_{t+h} \leq r_j \right\} - r_j \right) \right)^2 && \text{(A.19)}\\
&\leq \frac{1}{P} \left(\sum_{j=1}^{k} |\omega_j| \right)^2 < \infty . && \text{(A.20)}
\end{aligned}$$

This implies that the second moment of X_{t+h} is bounded.

The last condition to verify is a Lindeberg condition. We have to show that for every $\epsilon > 0$, as $P \to \infty$,

$$\frac{1}{\sigma^2_{S_P}} \sum_{t=E}^{T} \mathbb{E}\left[X^2_{t+h} \cdot \mathbb{1} \left\{ |X_{t+h}| > \epsilon \cdot \sigma_{S_P} \right\} \right] \longrightarrow 0 . \tag{A.21}$$

First, notice that $\sup_{P \geq 1} \left\{ \sigma_{S_P} \right\} < \infty$ from the absolute convergence of the sample covariance, and its limit is strictly positive by assumption.

Consider $|X_{t+h}|$ in the argument of the indicator function in Equation (A.21):

$$|X_{t+h}| = \left| \frac{1}{\sqrt{P}} \sum_{j=1}^{k} \omega_j \left(\mathbb{1}\left\{ \xi_{t+h} \leq r_j \right\} - r_j \right) \right| \tag{A.22}$$

$$\leq \frac{1}{\sqrt{P}} \sum_{j=1}^{k} |\omega_j|, \tag{A.23}$$

which can be made arbitrarily close to zero. Coupled with the previous properties of σ_{S_P} and the boundedness of X^2_{t+h}, this ensures that for any given $\varepsilon > 0$, the indicator function evaluates to zero for large enough P, hence the Lindeberg condition is satisfied.

This concludes verifying the conditions of Theorem 1.1 in Janson (2021), leading to the conclusion that $\frac{S_P}{\sigma_{S_P}} \xrightarrow{d} \mathcal{N}(0,1)$, hence proving the finite-dimensional convergence in distribution.

Step 3. By an argument analogous to Theorem 22.1 of Billingsley (1968), we can invoke Theorem 15.5 in Billingsley (1968), hence $\Psi_P(\cdot)$ is tight and $\mathbb{P}\left\{\Psi(\cdot) \in C[0,1]\right\} = 1$.

Notice that Theorem 22.1 of Billingsley (1968) assumes strict stationary and φ-mixing. In the proof, strict stationarity is only used in the application of Lemma 1 in Section 22. The lemma provides an upper bound on the fourth moment of the sum of strictly stationary random variables $\{W_i\}$ under the assumptions that $|W_0| \leq 1$ with probability 1, $\mathbb{E}\left(W_0\right) = 0$ and $\sum_{k=1}^{\infty} k^2 \varphi_k^{1/2} < \infty$. An applicable inequality of similar form, $\mathbb{E}\left[\left(\sum_{i=1}^{n} W_i\right)^4\right] \leq K_m \left(n^2 p + np\right)$, also holds under assuming the following: $|W_i| \leq 1$ with probability 1 for all i, $\mathbb{E}\left(W_i\right) = 0$ for all i, $\mathbb{E}\left(W_i^2\right) = p$ for all i, and m-dependence ($h-1$-dependence in our case). Strict stationarity is not required, and K_m is a constant that depends on m alone. This completes the proof. □

Proof of Theorem 2. The proof follows from a particular case of Theorem 2.3 in Inoue (2001) by setting (in Inoue's notation) $r = 1 - \delta = 1$ and $\hat{F}_n(t;\omega) = F(t) = t$. Assumption B (i) and (ii) in Inoue (2001) are satisfied by our Assumption 2 and Lemma 1, while Assumption B (iii) is irrelevant under the null hypothesis. Assumption C in Inoue (2001) holds by our Lemma 2, and we adopted his Assumption D as our Assumption 3. □

Appendix B Bootstrap simulation algorithm

This section describes the bootstrap simulation algorithm. The bootstrap procedure is only required for multi-step-ahead forecasts, *i.e.*, $h > 1$. For one-step-ahead forecasts, the critical values reported in Table 1 in the main text apply.

Let S denote the number of bootstrap samples. Recall that to compute the test statistics from the sample and their bootstrapped versions, one needs to discretize the unit interval (or a subset thereof if restricting attention to $\rho \in [0, 1]$). Following Rossi and Sekhposyan (2019), our choice is $[0 : 0.001 : 1]$, *i.e.*, discretizing the unit interval with a uniform stepsize of 0.001.

1. Calculate test statistics κ_P and C_p given by Equations (8) and (9) in the main text, respectively.
2. Simulate $\left\{ \{\eta_t^{(s)}\}_{t=E}^{T-\ell+1} \right\}_{s=1}^{S} \overset{iid}{\sim} \mathcal{N}(0, 1/\ell)$ and compute $\left\{ \Psi_P^{*(s)}(\cdot) \right\}_{s=1}^{S}$ based on Equation (6) in the main text.
3. Compute the bootstrap analogs of the test statistics, $\left\{ \kappa_P^{*(s)} \right\}_{s=1}^{S}$ and $\left\{ C_P^{*(s)} \right\}_{s=1}^{S}$, from $\left\{ \Psi_P^{*(s)}(\cdot) \right\}_{s=1}^{S}$.
4. Estimate critical values $\widehat{c}_{\kappa,P}$ and $\widehat{c}_{C,P}$ as $\alpha \cdot 100$ percentiles of the simulated $\left\{ \kappa_P^{*(s)} \right\}_{s=1}^{S}$ and $\left\{ C_P^{*(s)} \right\}_{s=1}^{S}$, respectively.
5. Reject the null hypothesis of correct calibration at the $\alpha \cdot 100\%$ significance level if $\kappa_P > \widehat{c}_{\kappa,P}$ or $C_p > \widehat{c}_{C,P}$, and fail to reject otherwise.

Appendix C Additional Monte Carlo simulation results

This section provides further Monte Carlo simulation results. See Section 4 for simulation design details.

Table C.1 shows the empirical rejection rates at the 5% nominal size for DGPs S1–S3, based on the asymptotic critical values for all sample sizes. That is, the difference between Table C.1 here and Table 2 reported in the main text is in the columns corresponding to $P = \{25, 50, 100, 200\}$, while the results for large samples (*i.e.*, $P = \{500, 1000\}$) are identical. As we can see, the use of finite-sample critical values leads to improved size in small samples. To confirm this numerically, for each DGP we computed the Frobenius norm of the matrix formed by the difference of the empirical rejection rates based on finite-sample critical values (11 rows and 8 columns corresponding to the KS and CvM statistics for $P = \{25, 50, 100, 200\}$) and a matrix with all values equal to 0.05, and repeated the procedure for empirical rejection rates based on asymptotic critical values. We label these matrices by the DGP and the type of critical values, *e.g.*, for DGP S1 and finite-sample critical values, we denote it by $\boldsymbol{D}_{\text{S1,finite}}$. The Frobenius norms are as follows:

$$\left|\boldsymbol{D}_{\text{S1,finite}}\right|_F = 0.021 < 0.038 = \left|\boldsymbol{D}_{\text{S1,asymptotic}}\right|_F, \tag{C.1}$$

$$\left|\boldsymbol{D}_{\text{S2,finite}}\right|_F = 0.022 < 0.037 = \left|\boldsymbol{D}_{\text{S2,asymptotic}}\right|_F, \tag{C.2}$$

$$\left|\boldsymbol{D}_{\text{S3,finite}}\right|_F = 0.026 < 0.044 = \left|\boldsymbol{D}_{\text{S3,asymptotic}}\right|_F, \tag{C.3}$$

verifying the improvements due to the finite-sample critical values.

Table C.1: Size properties based on asymptotic critical values: DGPs S1–S3

Panel A: DGP S1

Region or weight function		$P=25$		$P=50$		$P=100$		$P=200$		$P=500$		$P=1000$	
		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0,1]$	0.044	0.051	0.047	0.049	0.051	0.047	0.045	0.046	0.054	0.051	0.048	0.050
Left tail	$r \in [0,0.1]$	0.044	0.043	0.043	0.045	0.047	0.048	0.048	0.050	0.054	0.049	0.044	0.050
Right tail	$r \in [0.9,1]$	0.044	0.046	0.046	0.046	0.045	0.046	0.053	0.050	0.051	0.048	0.042	0.050
Left half	$r \in [0,0.5]$	0.045	0.050	0.048	0.047	0.048	0.047	0.047	0.046	0.049	0.053	0.051	0.051
Right half	$r \in [0.5,1]$	0.047	0.050	0.047	0.050	0.049	0.047	0.046	0.045	0.048	0.049	0.051	0.050
Center	$r \in [0.1,0.9]$	0.044	0.051	0.047	0.050	0.051	0.048	0.045	0.046	0.054	0.050	0.048	0.049
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.055	0.053	0.047	0.051	0.046	0.049	0.049	0.048	0.045	0.050	0.049	0.051
Left tail	$w(r)=(1-r)^2$	0.046	0.050	0.047	0.048	0.049	0.047	0.053	0.047	0.050	0.052	0.051	0.051
Right tail	$w(r)=r^2$	0.047	0.050	0.046	0.048	0.049	0.048	0.048	0.047	0.045	0.050	0.048	0.050
Center	$w(r)=r(1-r)$	0.043	0.052	0.046	0.049	0.049	0.049	0.045	0.046	0.051	0.050	0.049	0.048
Tails	$w(r)=(2r-1)^2$	0.070	0.050	0.059	0.046	0.051	0.050	0.052	0.047	0.050	0.049	0.052	0.052

Panel B: DGP S2

Region or weight function		$P=25$		$P=50$		$P=100$		$P=200$		$P=500$		$P=1000$	
		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0,1]$	0.048	0.051	0.046	0.047	0.047	0.049	0.051	0.051	0.052	0.053	0.048	0.046
Left tail	$r \in [0,0.1]$	0.040	0.042	0.048	0.053	0.046	0.048	0.044	0.046	0.051	0.048	0.041	0.047
Right tail	$r \in [0.9,1]$	0.044	0.043	0.047	0.048	0.046	0.047	0.052	0.049	0.051	0.048	0.043	0.049
Left half	$r \in [0,0.5]$	0.046	0.050	0.049	0.051	0.046	0.047	0.049	0.050	0.046	0.049	0.051	0.047
Right half	$r \in [0.5,1]$	0.050	0.053	0.046	0.047	0.049	0.052	0.052	0.054	0.048	0.052	0.052	0.051
Center	$r \in [0.1,0.9]$	0.047	0.050	0.046	0.046	0.047	0.050	0.051	0.051	0.052	0.053	0.048	0.046
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.051	0.049	0.050	0.053	0.048	0.047	0.046	0.048	0.045	0.049	0.049	0.049
Left tail	$w(r)=(1-r)^2$	0.043	0.050	0.048	0.050	0.047	0.047	0.048	0.049	0.048	0.051	0.045	0.046
Right tail	$w(r)=r^2$	0.048	0.053	0.048	0.048	0.051	0.051	0.054	0.053	0.050	0.052	0.051	0.051
Center	$w(r)=r(1-r)$	0.048	0.051	0.045	0.047	0.047	0.051	0.049	0.052	0.049	0.053	0.051	0.047
Tails	$w(r)=(2r-1)^2$	0.068	0.049	0.066	0.051	0.054	0.047	0.051	0.049	0.050	0.048	0.048	0.050

Panel C: DGP S3

Region or weight function		$P=25$		$P=50$		$P=100$		$P=200$		$P=500$		$P=1000$	
		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0,1]$	0.045	0.049	0.043	0.046	0.050	0.048	0.046	0.048	0.053	0.051	0.045	0.047
Left tail	$r \in [0,0.1]$	0.040	0.037	0.041	0.044	0.047	0.046	0.045	0.050	0.052	0.049	0.039	0.047
Right tail	$r \in [0.9,1]$	0.043	0.044	0.045	0.045	0.048	0.049	0.052	0.046	0.050	0.048	0.042	0.053
Left half	$r \in [0,0.5]$	0.042	0.047	0.044	0.044	0.048	0.046	0.049	0.047	0.049	0.053	0.048	0.047
Right half	$r \in [0.5,1]$	0.047	0.051	0.047	0.045	0.046	0.047	0.048	0.050	0.047	0.049	0.049	0.047
Center	$r \in [0.1,0.9]$	0.045	0.048	0.043	0.046	0.050	0.047	0.046	0.048	0.053	0.052	0.045	0.048
Tails	$r \in \{[0,0.1] \cup [0.9,1]\}$	0.052	0.049	0.042	0.046	0.049	0.048	0.047	0.047	0.045	0.050	0.043	0.049
Left tail	$w(r)=(1-r)^2$	0.042	0.049	0.045	0.046	0.049	0.045	0.047	0.049	0.049	0.053	0.048	0.045
Right tail	$w(r)=r^2$	0.048	0.051	0.048	0.047	0.048	0.047	0.049	0.049	0.048	0.049	0.049	0.049
Center	$w(r)=r(1-r)$	0.040	0.049	0.045	0.044	0.048	0.047	0.048	0.048	0.052	0.052	0.049	0.048
Tails	$w(r)=(2r-1)^2$	0.067	0.051	0.055	0.048	0.055	0.048	0.050	0.046	0.050	0.049	0.047	0.046

Note: The table reports the empirical rejection rates of the Kolmogorov–Smirnov- (KS) and Cramér–von Mises- (CvM) type tests in various regions and with various weights (in rows) for different sample sizes, denoted by P (in columns) at the 5% nominal size for DGPs S1–S3. For all sample sizes P, the asymptotic critical values reported in Table 1 are used. The number of Monte Carlo simulations is 10,000. The domain of r is discretized over $[0 : 0.001 : 1]$.

Table C.2 is the counterpart of Table 3 (Panels B to D) in the main text, with the difference that here the bootstrap block size is set as $\ell = \lfloor P^{1/3} \rfloor$, instead of our proposed scheme $\ell = \max\{h-1, \lfloor P^{1/3} \rfloor\}$. In our setup, this leads to different block sizes when $h = 4$ and $P = 25$, and for all sample sizes when $h = 12$. While the differences in rejection rates are quite small, our proposed scheme consistently outperforms the one based only on the size of the out-of-sample period P. This is verified by comparing the Frobenius norms of the matrices formed by taking the difference between the empirical rejection rates and a matrix with all values equal to 0.05. Labeling such matrices by h, P, M (standing for the max operator of the proposed rule) and O (denoting the rule based on P only), we obtained the following values:

$$|\boldsymbol{D}^{M}_{h=4,P=25}|_F = 0.169 < 0.176 = |\boldsymbol{D}^{O}_{h=4,P=25}|_F, \tag{C.4}$$

$$|\boldsymbol{D}^{M}_{h=12,P=25}|_F = 0.073 < 0.180 = |\boldsymbol{D}^{O}_{h=12,P=25}|_F, \tag{C.5}$$

$$|\boldsymbol{D}^{M}_{h=12,P=50}|_F = 0.096 < 0.124 = |\boldsymbol{D}^{O}_{h=12,P=50}|_F, \tag{C.6}$$

$$|\boldsymbol{D}^{M}_{h=12,P=100}|_F = 0.076 < 0.100 = |\boldsymbol{D}^{O}_{h=12,P=100}|_F, \tag{C.7}$$

$$|\boldsymbol{D}^{M}_{h=12,P=200}|_F = 0.060 < 0.085 = |\boldsymbol{D}^{O}_{h=12,P=200}|_F, \tag{C.8}$$

$$|\boldsymbol{D}^{M}_{h=12,P=500}|_F = 0.042 < 0.059 = |\boldsymbol{D}^{O}_{h=12,P=500}|_F, \tag{C.9}$$

$$|\boldsymbol{D}^{M}_{h=12,P=1000}|_F = 0.047 < 0.055 = |\boldsymbol{D}^{O}_{h=12,P=1000}|_F, \tag{C.10}$$

confirming that the proposed rule leads to rejection rates which are slightly closer to the nominal rate than the alternative rule.

Table C.2: Size properties based on bootstrap block length $\ell = \lfloor P^{1/3} \rfloor$: DGP S5

Panel A: DGP S5, $h = 2$

Region or weight function		$P = 25$		$P = 50$		$P = 100$		$P = 200$		$P = 500$		$P = 1000$	
		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0, 1]$	0.025	0.018	0.035	0.026	0.036	0.027	0.041	0.031	0.045	0.038	0.043	0.043
Left tail	$r \in [0, 0.1]$	0.094	0.127	0.085	0.086	0.070	0.062	0.061	0.055	0.053	0.049	0.051	0.051
Right tail	$r \in [0.9, 1]$	0.099	0.134	0.083	0.080	0.070	0.061	0.062	0.058	0.052	0.050	0.054	0.050
Left half	$r \in [0, 0.5]$	0.029	0.022	0.038	0.030	0.041	0.033	0.042	0.037	0.046	0.042	0.043	0.043
Right half	$r \in [0.5, 1]$	0.032	0.025	0.036	0.029	0.039	0.030	0.042	0.034	0.043	0.038	0.043	0.043
Center	$r \in [0.1, 0.9]$	0.025	0.018	0.035	0.026	0.036	0.028	0.041	0.031	0.045	0.037	0.043	0.042
Tails	$r \in \{[0, 0.1] \cup [0.9, 1]\}$	0.049	0.062	0.054	0.055	0.050	0.050	0.052	0.049	0.048	0.048	0.051	0.047
Left tail	$w(r) = (1-r)^2$	0.037	0.021	0.044	0.030	0.046	0.032	0.050	0.036	0.047	0.042	0.051	0.044
Right tail	$w(r) = r^2$	0.040	0.024	0.047	0.031	0.042	0.029	0.048	0.033	0.046	0.035	0.049	0.045
Center	$w(r) = r(1-r)$	0.029	0.019	0.035	0.025	0.038	0.028	0.041	0.032	0.045	0.038	0.042	0.042
Tails	$w(r) = (2r-1)^2$	0.013	0.027	0.028	0.032	0.036	0.034	0.044	0.038	0.043	0.039	0.051	0.045

Panel B: DGP S5, $h = 4$

Region or weight function		$P = 25$		$P = 50$		$P = 100$		$P = 200$		$P = 500$		$P = 1000$	
		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0, 1]$	0.016	0.008	0.024	0.015	0.029	0.018	0.034	0.026	0.035	0.028	0.040	0.036
Left tail	$r \in [0, 0.1]$	0.084	0.117	0.081	0.079	0.061	0.057	0.057	0.050	0.050	0.048	0.050	0.047
Right tail	$r \in [0.9, 1]$	0.091	0.121	0.076	0.074	0.066	0.056	0.056	0.049	0.049	0.045	0.054	0.050
Left half	$r \in [0, 0.5]$	0.021	0.015	0.029	0.020	0.032	0.025	0.038	0.031	0.035	0.031	0.040	0.038
Right half	$r \in [0.5, 1]$	0.021	0.013	0.028	0.020	0.031	0.025	0.036	0.032	0.036	0.032	0.040	0.037
Center	$r \in [0.1, 0.9]$	0.017	0.009	0.024	0.015	0.029	0.018	0.034	0.026	0.035	0.028	0.040	0.037
Tails	$r \in \{[0, 0.1] \cup [0.9, 1]\}$	0.045	0.059	0.051	0.050	0.043	0.044	0.047	0.042	0.044	0.045	0.050	0.050
Left tail	$w(r) = (1-r)^2$	0.028	0.014	0.039	0.019	0.040	0.024	0.043	0.029	0.039	0.031	0.043	0.039
Right tail	$w(r) = r^2$	0.030	0.014	0.038	0.019	0.040	0.023	0.042	0.029	0.041	0.032	0.044	0.037
Center	$w(r) = r(1-r)$	0.019	0.008	0.026	0.016	0.031	0.018	0.035	0.027	0.036	0.029	0.042	0.037
Tails	$w(r) = (2r-1)^2$	0.012	0.019	0.025	0.025	0.030	0.026	0.039	0.032	0.040	0.035	0.045	0.040

Panel C: DGP S5, $h = 12$

Region or weight function		$P = 25$		$P = 50$		$P = 100$		$P = 200$		$P = 500$		$P = 1000$	
		KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM	KS	CvM
Full	$r \in [0, 1]$	0.016	0.011	0.023	0.015	0.027	0.019	0.031	0.021	0.035	0.030	0.037	0.033
Left tail	$r \in [0, 0.1]$	0.090	0.124	0.078	0.074	0.060	0.054	0.058	0.054	0.051	0.051	0.047	0.049
Right tail	$r \in [0.9, 1]$	0.094	0.128	0.080	0.077	0.072	0.061	0.058	0.054	0.049	0.046	0.046	0.048
Left half	$r \in [0, 0.5]$	0.021	0.015	0.024	0.018	0.031	0.024	0.034	0.025	0.037	0.034	0.039	0.036
Right half	$r \in [0.5, 1]$	0.022	0.017	0.029	0.021	0.030	0.024	0.033	0.028	0.040	0.035	0.037	0.035
Center	$r \in [0.1, 0.9]$	0.017	0.011	0.023	0.014	0.027	0.019	0.031	0.022	0.035	0.030	0.037	0.033
Tails	$r \in \{[0, 0.1] \cup [0.9, 1]\}$	0.046	0.060	0.049	0.051	0.048	0.045	0.049	0.048	0.047	0.045	0.043	0.047
Left tail	$w(r) = (1-r)^2$	0.030	0.016	0.033	0.018	0.034	0.023	0.042	0.024	0.044	0.034	0.043	0.036
Right tail	$w(r) = r^2$	0.032	0.017	0.037	0.021	0.040	0.023	0.037	0.025	0.044	0.034	0.041	0.034
Center	$w(r) = r(1-r)$	0.020	0.010	0.027	0.015	0.029	0.018	0.033	0.022	0.035	0.031	0.039	0.034
Tails	$w(r) = (2r-1)^2$	0.012	0.021	0.024	0.025	0.033	0.025	0.041	0.030	0.043	0.036	0.043	0.036

Note: The table reports the empirical rejection rates of the Kolmogorov–Smirnov- (KS) and Cramér–von Mises- (CvM) type tests in various regions and with various weights (in rows) for different sample sizes, denoted by P (in columns) at the 5% nominal size for DGP S5. For all sample sizes P, the critical values are based on 5000 bootstrap replications with block length $\ell = \lfloor P^{1/3} \rfloor$. The number of Monte Carlo simulations is 10,000. The domain of r is discretized over $[0 : 0.001 : 1]$.

Appendix D Estimation details of the UC and UCSV models

This appendix describes the prior specifications and posterior samplers used to estimate the **UC** and **UCSV** models. Each model is recursively re-estimated at each forecast origin. In what follows, T denotes the sample size at each forecast origin. After 2,000 burn-in draws, for each of the next 10,000 draws from the posterior we simulate one path from the predictive distribution.

Let $\boldsymbol{y} \equiv (y_1, \ldots, y_T)'$ denote the vector of observations, and define similarly the vector of unobserved trend components $\boldsymbol{\tau} \equiv (\tau_1, \ldots, \tau_T)'$. Define the $(T \times T)$ first difference matrix as

$$\boldsymbol{H} \equiv \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}, \tag{D.1}$$

let $\mathbf{1}_T$ denote a $(T \times 1)$ vector of ones, and let $\boldsymbol{I}_T$ stand for the $(T \times T)$ identity matrix.

D.1 UC model

The prior distribution on the initial value of the trend component is normal: $\tau_0 \sim \mathcal{N}(\underline{\mu}_{\tau_0}, \underline{\sigma}^2_{\tau_0})$, with $\underline{\mu}_{\tau_0} = 0$ and $\underline{\sigma}^2_{\tau_0} = 5^2$ for both the monthly and the quarterly models. The prior on the variance of the shocks to the trend component is inverse gamma: $\sigma^2_\tau \sim \mathcal{IG}(\underline{\nu}_{\sigma^2_\tau}, \underline{S}_{\sigma^2_\tau})$, with $\underline{\nu}_{\sigma^2_\tau} = 3$ for both the monthly and the quarterly models, and $\underline{S}_{\sigma^2_\tau} = (\underline{\nu}_{\sigma^2_\tau} - 1)0.2^2$ in the case of the monthly model, and $\underline{S}_{\sigma^2_\tau} = (\underline{\nu}_{\sigma^2_\tau} - 1)0.5^2$ in the case of the quarterly model. Finally, the prior on the variance of the shocks to the cyclical component is inverse gamma: $\sigma^2_y \sim \mathcal{IG}(\underline{\nu}_{\sigma^2_y}, \underline{S}_{\sigma^2_y})$, with $\underline{\nu}_{\sigma^2_y} = 3$ for both the monthly and the quarterly models, and $\underline{S}_{\sigma^2_y} = (\underline{\nu}_{\sigma^2_y} - 1)5^2$ in the case of the monthly model, and $\underline{S}_{\sigma^2_y} = (\underline{\nu}_{\sigma^2_y} - 1)8^2$ in the case of the quarterly model. All the priors are independent.

After initialization, the Gibbs sampler cycles over the following steps drawing from the conditional posteriors (suppressing the sweep index for simplicity):

1. Draw $\left(\boldsymbol{\tau} \mid \boldsymbol{y}, \sigma_y^2, \sigma_\tau^2, \tau_0\right) \sim \mathcal{N}\left(\overline{\boldsymbol{\mu}}_{\boldsymbol{\tau}}, \overline{\boldsymbol{\Sigma}}_{\boldsymbol{\tau}}^{-1}\right)$, where

$$\overline{\boldsymbol{\mu}}_{\boldsymbol{\tau}} = \overline{\boldsymbol{\Sigma}}_{\boldsymbol{\tau}}^{-1} \left(\frac{\tau_0}{\sigma_\tau^2} \boldsymbol{H}'\boldsymbol{H}\boldsymbol{1}_T + \frac{1}{\sigma_y^2}\boldsymbol{y} \right), \tag{D.2}$$

$$\overline{\boldsymbol{\Sigma}}_{\boldsymbol{\tau}} = \frac{1}{\sigma_\tau^2}\boldsymbol{H}'\boldsymbol{H} + \frac{1}{\sigma_y^2}\boldsymbol{I}_T, \tag{D.3}$$

based on the precision sampler proposed by Chan and Jeliazkov (2009).

2. Draw $\left(\sigma_y^2 \mid \boldsymbol{y}, \boldsymbol{\tau}, \sigma_\tau^2, \tau_0\right) \sim \mathcal{IG}(\overline{\nu}_{\sigma_y^2}, \overline{S}_{\sigma_y^2})$, where

$$\overline{\nu}_{\sigma_y^2} = \underline{\nu}_{\sigma_y^2} + \frac{T}{2}, \tag{D.4}$$

$$\overline{S}_{\sigma_y^2} = \underline{S}_{\sigma_y^2} + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{\tau})'(\boldsymbol{y} - \boldsymbol{\tau}). \tag{D.5}$$

3. Draw $\left(\sigma_\tau^2 \mid \boldsymbol{y}, \boldsymbol{\tau}, \sigma_y^2, \tau_0\right) \sim \mathcal{IG}(\overline{\nu}_{\sigma_\tau^2}, \overline{S}_{\sigma_\tau^2})$, where

$$\overline{\nu}_{\sigma_\tau^2} = \underline{\nu}_{\sigma_\tau^2} + \frac{T}{2}, \tag{D.6}$$

$$\overline{S}_{\sigma_\tau^2} = \underline{S}_{\sigma_\tau^2} + \frac{1}{2}(\boldsymbol{\tau} - \boldsymbol{1}_T\tau_0)'\boldsymbol{H}'\boldsymbol{H}(\boldsymbol{\tau} - \boldsymbol{1}_T\tau_0). \tag{D.7}$$

4. Draw $\left(\tau_0 \mid \boldsymbol{y}, \boldsymbol{\tau}, \sigma_y^2, \sigma_\tau^2\right) \sim \mathcal{N}\left(\overline{\mu}_{\tau_0}, \overline{\sigma}_{\tau_0}^2\right)$, where

$$\overline{\mu}_{\tau_0} = \overline{\sigma}_{\tau_0}^2 \left(\frac{\underline{\mu}_{\tau_0}}{\underline{\sigma}_{\tau_0}^2} + \frac{\tau_1}{\sigma_\tau^2} \right), \tag{D.8}$$

$$\overline{\sigma}_{\tau_0}^2 = \left(\frac{1}{\underline{\sigma}_{\tau_0}^2} + \frac{1}{\sigma_\tau^2} \right)^{-1}. \tag{D.9}$$

D.2 UCSV model

The priors on τ_0 and σ_τ^2 are identical to those in the **UC** model. The prior on the mean of the log volatility process is normal: $\mu_h \sim \mathcal{N}(\underline{\mu}_{\mu_h}, \underline{\sigma}_{\mu_h}^2)$, with $\underline{\mu}_{\mu_h} = 5$ and $\underline{\mu}_{\mu_h} = 3$ in the case of the monthly and quarterly models, respectively, while $\underline{\sigma}_{\mu_h}^2 = 5$ in both cases. The prior on the autoregressive coefficient is truncated normal: $\phi_h \sim \mathcal{N}(\underline{\mu}_{\phi_h}, \underline{\sigma}_{\phi_h}^2) \cdot \mathbb{1}(|\phi_h| < 1)$, with $\underline{\mu}_{\phi_h} = 0.9$ and $\underline{\sigma}_{\phi_h}^2 = 0.5^2$ for both the monthly and the quarterly models. The prior on the shocks to the log volatility process is inverse gamma: $\sigma_h^2 \sim \mathcal{IG}(\underline{\nu}_{\sigma_h^2}, \underline{S}_{\sigma_h^2})$, with $\underline{\nu}_{\sigma_h^2} = 3$ for both the monthly and the quarterly models, and $\underline{S}_{\sigma_h^2} = (\underline{\nu}_{\sigma_h^2} - 1)0.1^2$ in the case of the monthly model, and $\underline{S}_{\sigma_h^2} = (\underline{\nu}_{\sigma_h^2} - 1)0.2^2$ in the case of the quarterly model. All the priors are independent. The specification is completed by assuming that the initial log volatility is drawn from its stationary distribution: $h_1 \sim \mathcal{N}\left(\mu_h, \sigma_h^2/(1 - \phi_h^2)\right)$.

Collect the log volatility process in $\boldsymbol{h} \equiv (h_1, \ldots, h_T)'$ and define the diagonal covariance matrix of the cyclical component as $\boldsymbol{\Sigma}_{\varepsilon^y} \equiv \operatorname{diag}(\exp(h_1), \ldots, \exp(h_T))$.

After initialization, the Gibbs sampler cycles over the following steps drawing from the conditional posteriors (suppressing the sweep index for simplicity):

1. Draw $\left(\boldsymbol{\tau} \mid \boldsymbol{y}, \tau_0, \sigma_\tau^2, \boldsymbol{h}, \phi_h, \sigma_h^2, \mu_h\right) \sim \mathcal{N}\left(\overline{\boldsymbol{\mu}}_{\boldsymbol{\tau}}, \overline{\boldsymbol{\Sigma}}_{\boldsymbol{\tau}}^{-1}\right)$, where

$$\overline{\boldsymbol{\mu}}_{\boldsymbol{\tau}} = \overline{\boldsymbol{\Sigma}}_{\boldsymbol{\tau}}^{-1}\left(\frac{\tau_0}{\sigma_\tau^2}\boldsymbol{H}'\boldsymbol{H}\boldsymbol{1}_T + \boldsymbol{\Sigma}_{\varepsilon^y}^{-1}\boldsymbol{y}\right), \tag{D.10}$$

$$\overline{\boldsymbol{\Sigma}}_{\boldsymbol{\tau}} = \left[\sigma_\tau^2\left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1}\right]^{-1} + \boldsymbol{\Sigma}_{\varepsilon^y}^{-1}, \tag{D.11}$$

   based on the precision sampler proposed by Chan and Jeliazkov (2009).

2. Draw $\left(\tau_0 \mid \boldsymbol{y}, \boldsymbol{\tau}, \sigma_\tau^2, \boldsymbol{h}, \phi_h, \sigma_h^2, \mu_h\right) \sim \mathcal{N}\left(\overline{\mu}_{\tau_0}, \overline{\sigma}_{\tau_0}^2\right)$, where

$$\overline{\mu}_{\tau_0} = \overline{\sigma}_{\tau_0}^2\left(\frac{\underline{\mu}_{\tau_0}}{\underline{\sigma}_{\tau_0}^2} + \frac{\tau_1}{\sigma_\tau^2}\right), \tag{D.12}$$

$$\overline{\sigma}_{\tau_0}^2 = \left(\frac{1}{\sigma_\tau^2} + \frac{1}{\underline{\sigma}_{\tau_0}^2}\right)^{-1}. \tag{D.13}$$

3. Draw $\left(\sigma_\tau^2 \mid \boldsymbol{y}, \boldsymbol{\tau}, \tau_0, \boldsymbol{h}, \phi_h, \sigma_h^2, \mu_h\right) \sim \mathcal{IG}(\overline{\nu}_{\sigma_\tau^2}, \overline{S}_{\sigma_\tau^2})$, where

$$\overline{\nu}_{\sigma_\tau^2} = \underline{\nu}_{\sigma_\tau^2} + \frac{T}{2}, \tag{D.14}$$

$$\overline{S}_{\sigma_\tau^2} = \underline{S}_{\sigma_\tau^2} + \frac{1}{2}(\boldsymbol{\tau} - \boldsymbol{1}_T\tau_0)'\boldsymbol{H}'\boldsymbol{H}(\boldsymbol{\tau} - \boldsymbol{1}_T\tau_0). \tag{D.15}$$

4. Draw $\left(\boldsymbol{h} \mid \boldsymbol{y}, \boldsymbol{\tau}, \tau_0, \sigma_\tau^2, \phi_h, \sigma_h^2, \mu_h\right)$ using the precision sampler proposed by Chan and Hsiao (2014), with the 10-component mixture approximation to the $\log(\chi_1^2)$ density due to Omori et al. (2007).

5. Draw $\left(\phi_h \mid \boldsymbol{y}, \boldsymbol{\tau}, \tau_0, \sigma_\tau^2, \boldsymbol{h}, \sigma_h^2, \mu_h\right)$ in an Independent Metropolis–Hastings step, as in Chan and Hsiao (2014).

6. Draw $\left(\sigma_h^2 \mid \boldsymbol{y}, \boldsymbol{\tau}, \tau_0, \sigma_\tau^2, \boldsymbol{h}, \phi_h, \mu_h\right) \sim \mathcal{IG}(\overline{\nu}_{\sigma_h^2}, \overline{S}_{\sigma_h^2})$, where

$$\overline{\nu}_{\sigma_h^2} = \underline{\nu}_{\sigma_h^2} + \frac{T}{2}, \tag{D.16}$$

$$\overline{S}_{\sigma_h^2} = \underline{S}_{\sigma_h^2} + \frac{1}{2}\left((1-\phi_h^2)(h_1-\mu_h)^2 + \sum_{t=2}^{T}(h_t - \mu_h - \phi_h(h_{t-1}-\mu_h))^2\right), \tag{D.17}$$

   as in Chan and Hsiao (2014).

7. Draw $\left(\mu_h \mid \boldsymbol{y}, \boldsymbol{\tau}, \tau_0, \sigma^2_\tau, \boldsymbol{h}, \phi_h, \sigma^2_h\right) \sim \mathcal{N}(\overline{\mu}_{\mu_h}, \overline{\sigma}^2_{\mu_h})$, where

$$\overline{\mu}_{\mu_h} = \overline{\sigma}^2_{\mu_h} \left(\frac{\underline{\mu}_{\mu_h}}{\underline{\sigma}^2_{\mu_h}} + \boldsymbol{x}'_{\mu_h} \boldsymbol{\Sigma}_{\boldsymbol{h}}^{-1} \boldsymbol{z}_{\mu_h} \right), \tag{D.18}$$

$$\overline{\sigma}^2_{\mu_h} = \left(\frac{1}{\underline{\sigma}^2_{\mu_h}} + \boldsymbol{x}'_{\mu_h} \boldsymbol{\Sigma}_{\boldsymbol{h}}^{-1} \boldsymbol{x}_{\mu_h} \right)^{-1}, \tag{D.19}$$

$$\boldsymbol{x}_{\mu_h} = (1, 1-\phi_h, 1-\phi_h, \ldots, 1-\phi_h)', \tag{D.20}$$

$$\boldsymbol{z}_{\mu_h} = (h_1, h_2 - \phi_h h_1, \ldots, h_T - \phi_h h_{T-1})', \tag{D.21}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{h}} = \text{diag}\left(\frac{\sigma^2_h}{1-\phi^2_h}, \sigma^2_h, \ldots, \sigma^2_h \right), \tag{D.22}$$

following Chan and Hsiao (2014).

Appendix E Further empirical results

This section presents supplementary empirical results (predictive bands, empirical CDFs of PITs, and test outcomes) across the full set of forecast horizons.

Figure E.1 shows the predictive bands of the monthly frequency **UC** and **UCSV** models at a complete set of monthly horizons (between 1 and 12 months ahead), while Figure E.2 displays analogous results for one-year-ahead forecasts.

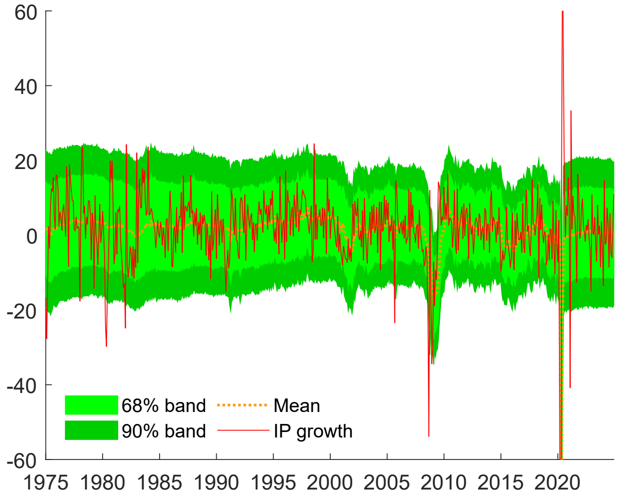
Figure E.3 shows the predictive bands of the quarterly frequency **UC** and **UCSV** models at a complete set of quarterly horizons (between 1 and 8 quarters ahead), while Figure E.4 displays analogous results for one- and two-year-ahead forecasts.

Figure E.5 shows the empirical CDFs and 5% critical value bands of PITs of the monthly frequency **UC** and **UCSV** models at a complete set of monthly horizons, while Figure E.6 shows them at the one-year-ahead horizon.

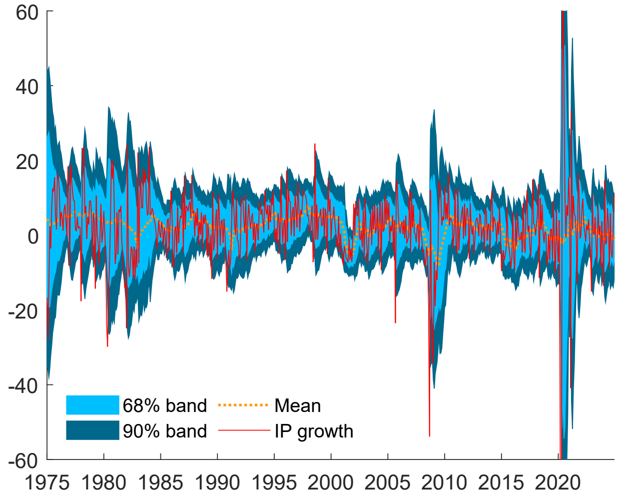
Figure E.7 shows the empirical CDFs and 5% critical value bands of PITs of the quarterly frequency **UC** and **UCSV** models at a complete set of quarterly horizons, while Figure E.8 displays the one- and two-year-ahead results.

For the monthly frequency models, Table E.1 shows the *p*-values of the tests of correct specification for the full set of monthly and annual horizons, while Table E.2 displays analogous results based on the quarterly models.

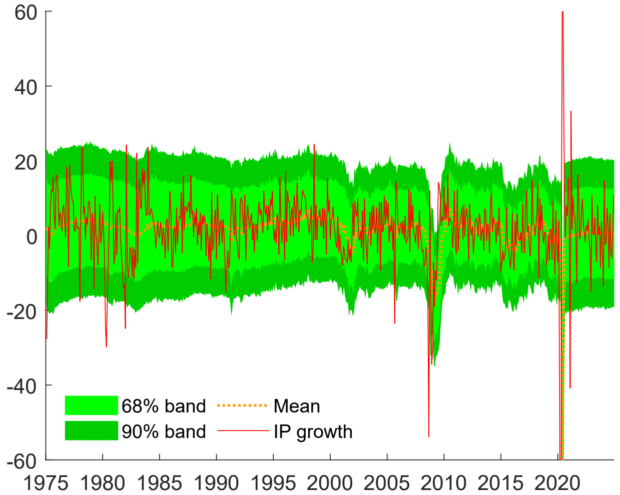
Figure E.1: Equal-tailed 68% and 90% predictive bands



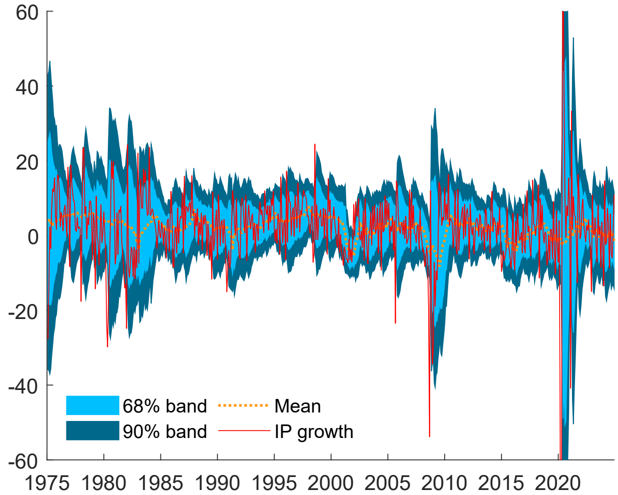
(a) **UC** model, $h = 1\text{M}$



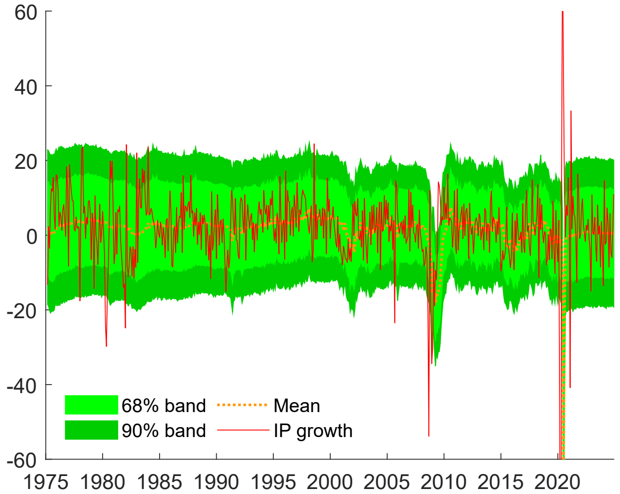
(b) **UCSV** model, $h = 1\text{M}$



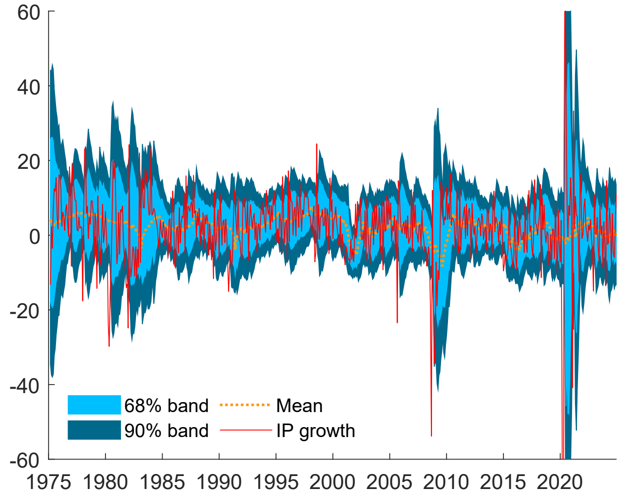
(c) **UC** model, $h = 2\text{M}$



(d) **UCSV** model, $h = 2\text{M}$

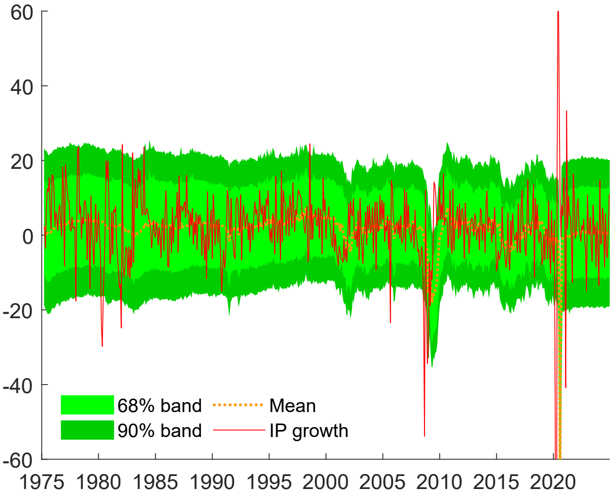


(e) **UC** model, $h = 3\text{M}$

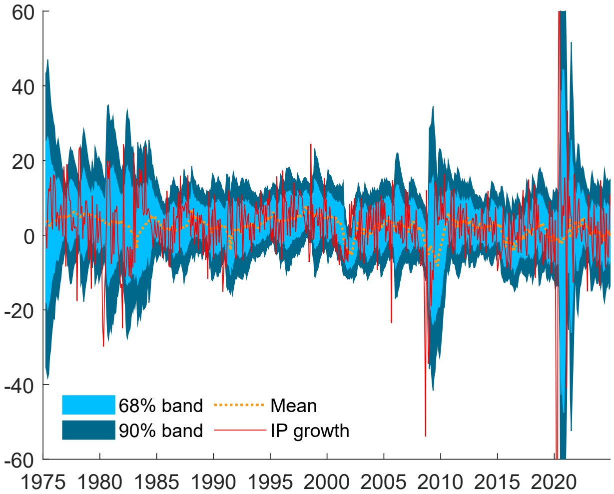


(f) **UCSV** model, $h = 3\text{M}$

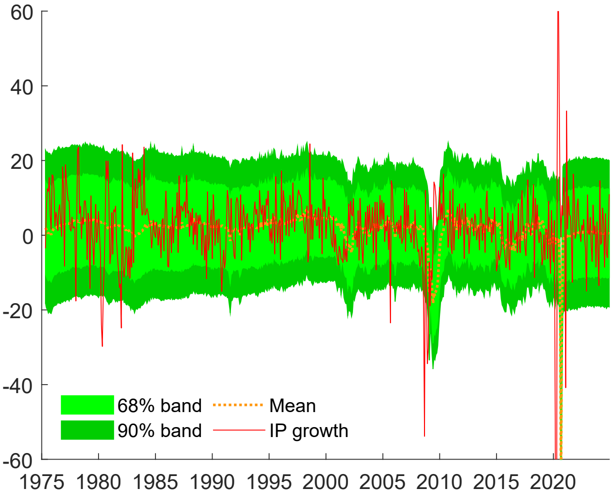
Figure E.1 (cont.): Equal-tailed 68% and 90% predictive bands



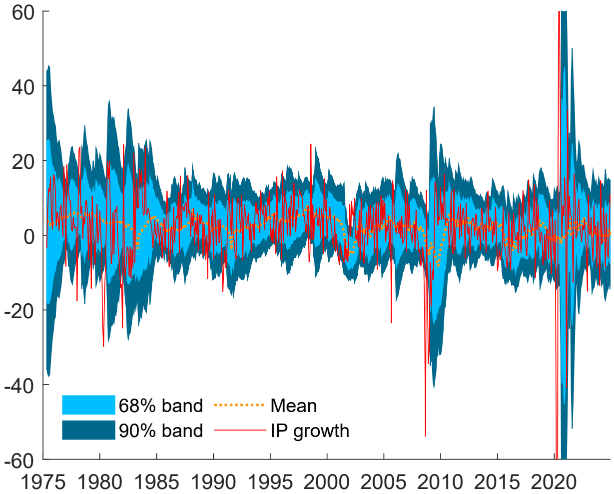
(g) **UC** model, $h = 4M$



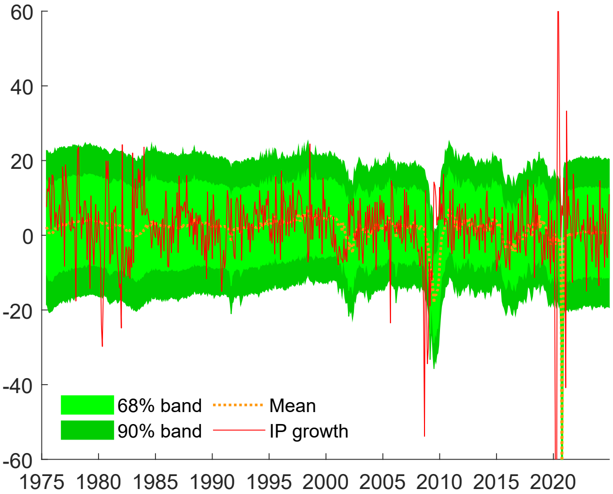
(h) **UCSV** model, $h = 4M$



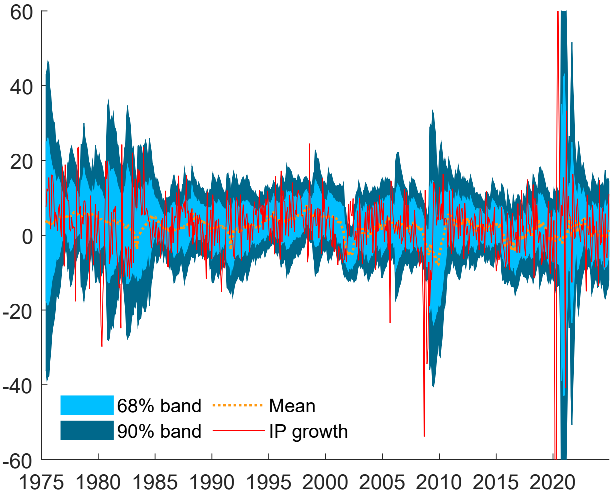
(i) **UC** model, $h = 5M$



(j) **UCSV** model, $h = 5M$

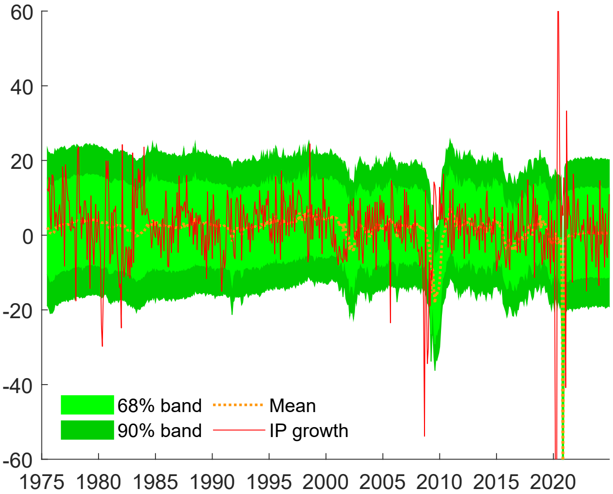


(k) **UC** model, $h = 6M$

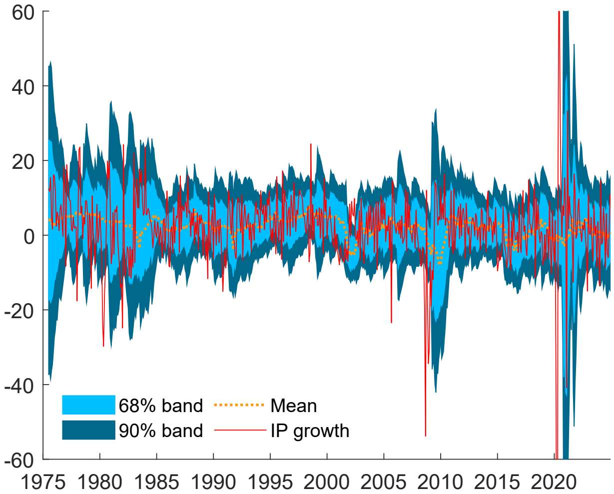


(l) **UCSV** model, $h = 6M$

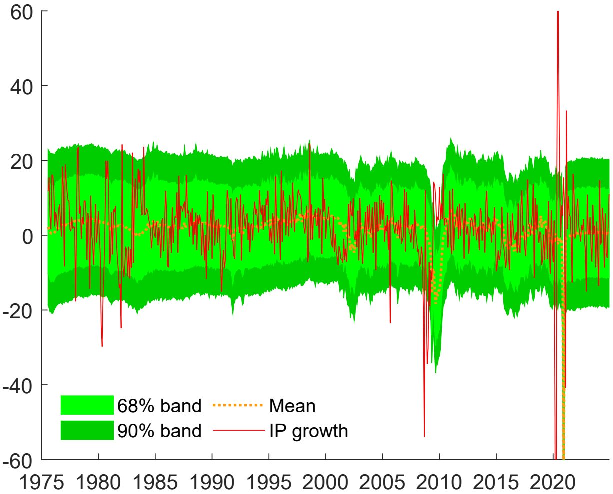
Figure E.1 (cont.): Equal-tailed 68% and 90% predictive bands



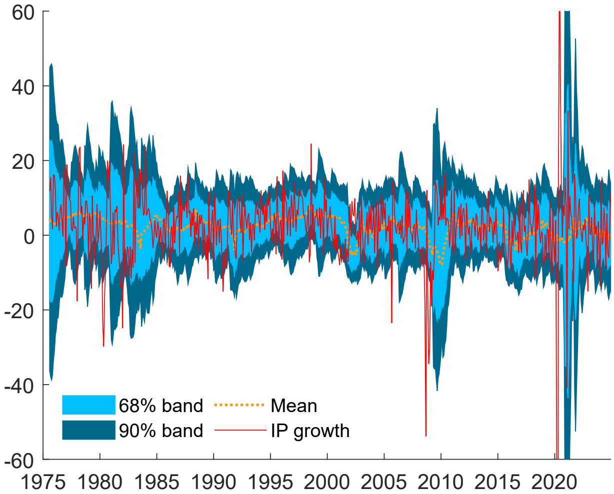
(m) **UC** model, $h = 7$M



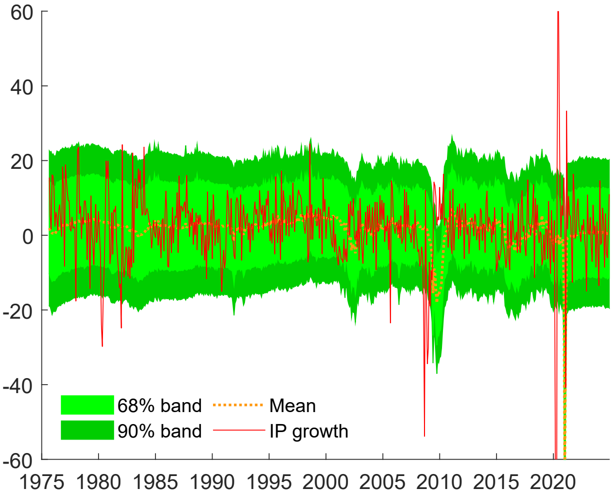
(n) **UCSV** model, $h = 7$M



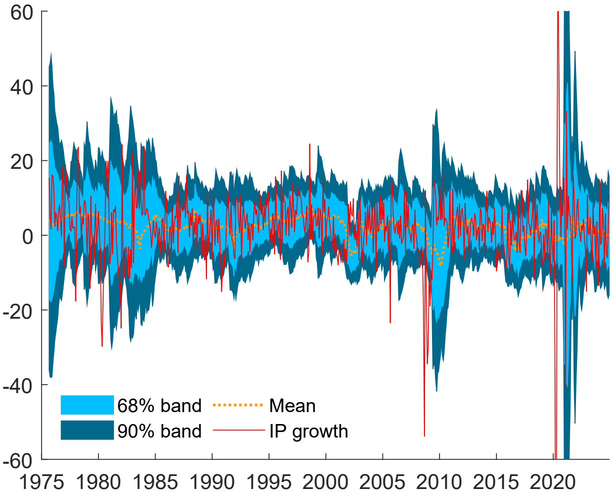
(o) **UC** model, $h = 8$M



(p) **UCSV** model, $h = 8$M

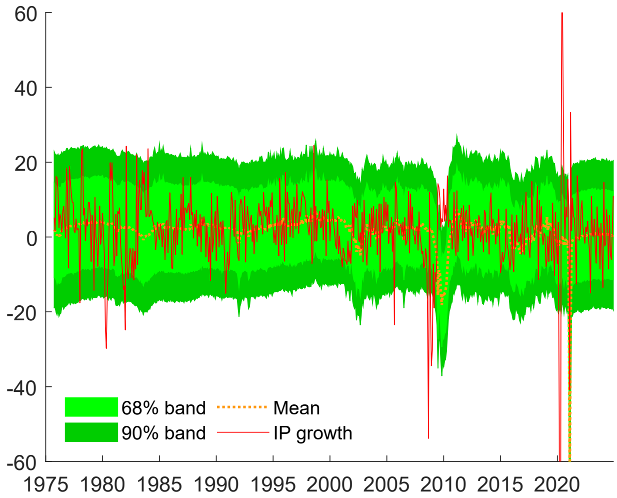


(q) **UC** model, $h = 9$M

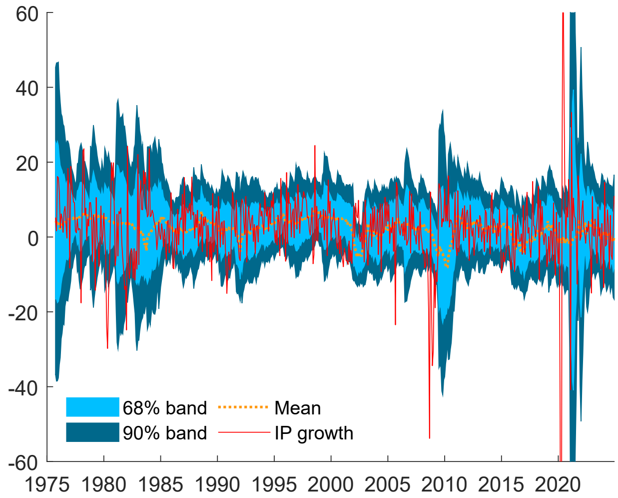


(r) **UCSV** model, $h = 9$M

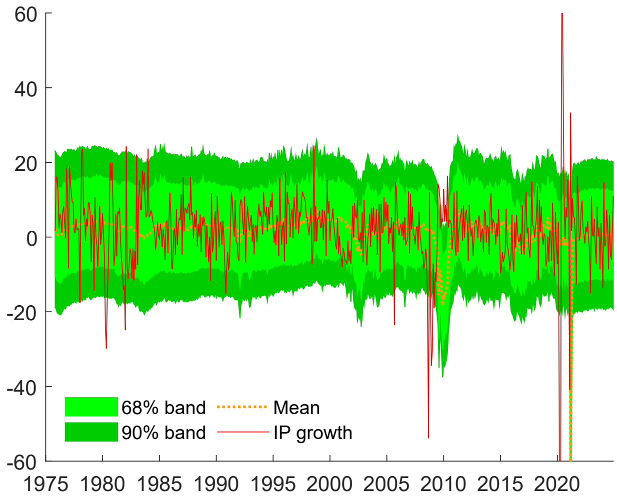
Figure E.1 (cont.): Equal-tailed 68% and 90% predictive bands



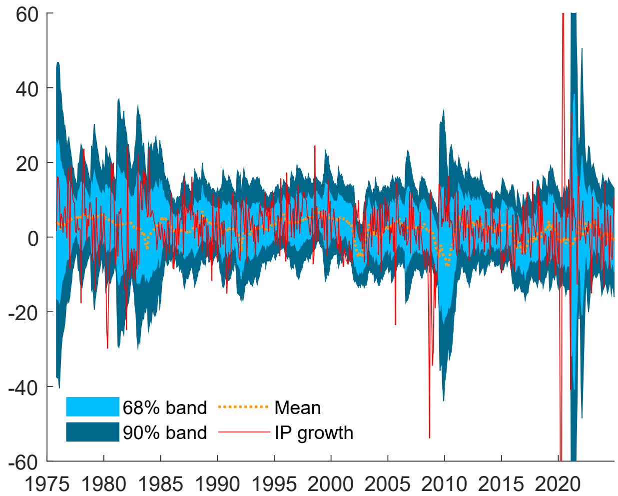
(s) **UC** model, $h = 10\text{M}$



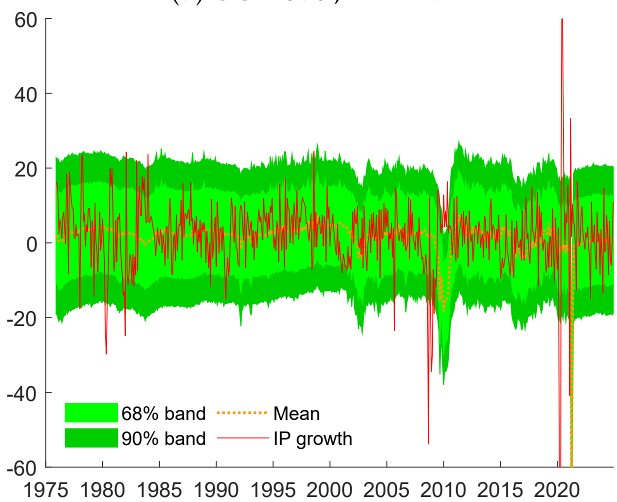
(t) **UCSV** model, $h = 10\text{M}$



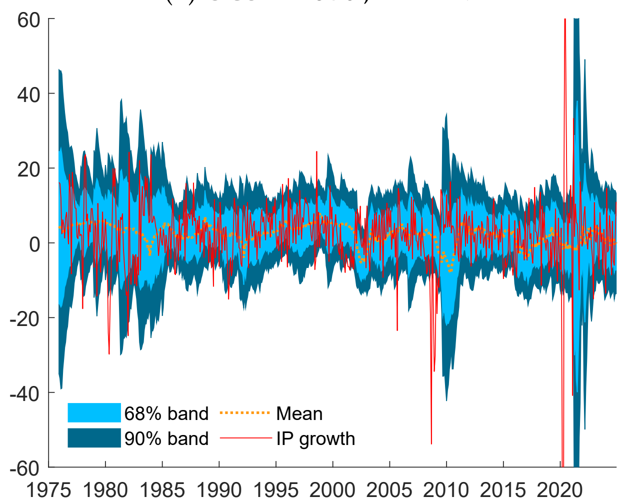
(u) **UC** model, $h = 11\text{M}$



(v) **UCSV** model, $h = 11\text{M}$



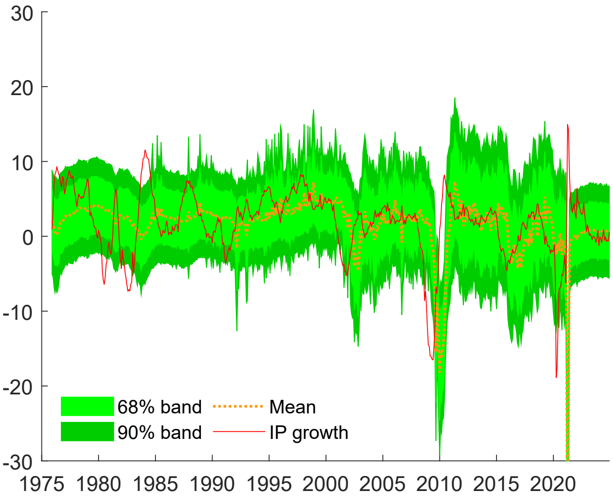
(w) **UC** model, $h = 12\text{M}$



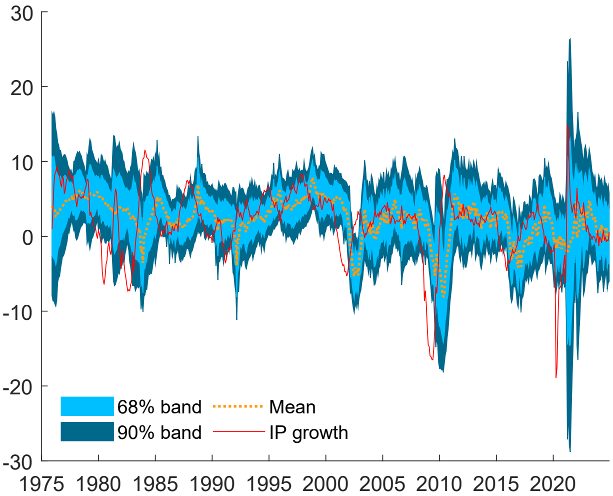
(x) **UCSV** model, $h = 12\text{M}$

Note: The panels show the 68% and 90% predictive bands (with light and dark shades, respectively) along with the mean predictions (dotted lines) and the realized annualized month-on-month industrial production growth (solid lines) between 1974:M12 + h and 2024:M12. Dates on the horizontal axes correspond to the forecast target dates.

Figure E.2: Equal-tailed 68% and 90% predictive bands



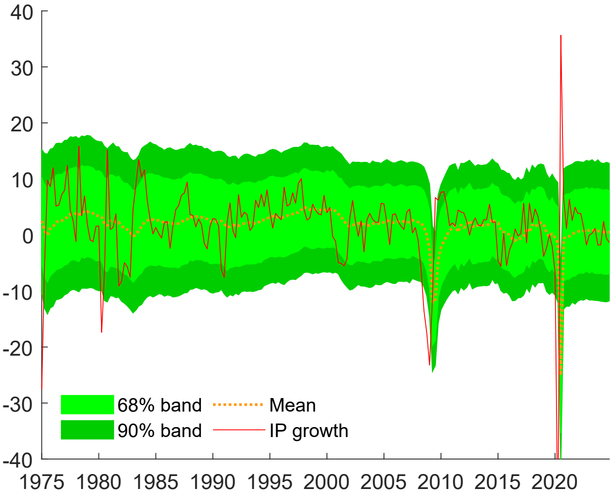
(a) **UC** model, $h = 1Y$



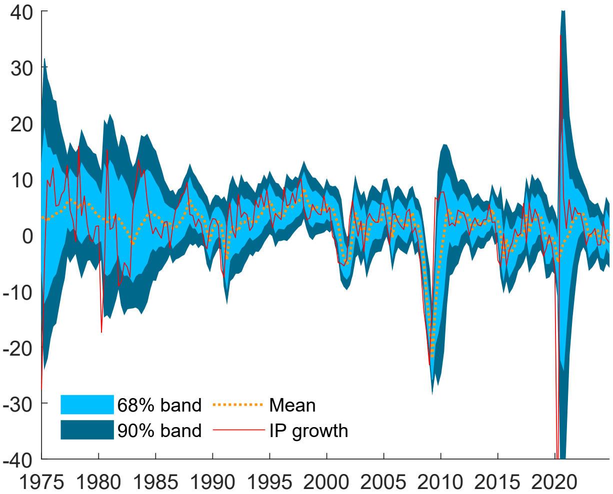
(b) **UCSV** model, $h = 1Y$

Note: The panels show the 68% and 90% predictive bands (with light and dark shades, respectively) along with the mean predictions (dotted lines) and the realized year-on-year industrial production growth (solid lines) between 1974:M12 + h and 2024:M12. Dates on the horizontal axes correspond to the forecast target dates.

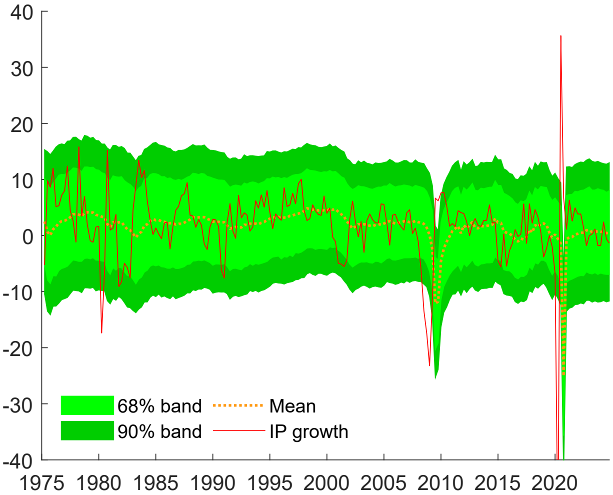
Figure E.3: Equal-tailed 68% and 90% predictive bands



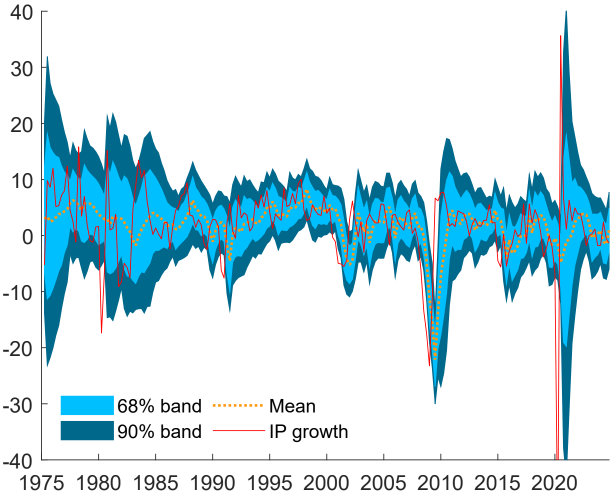
(a) **UC** model, $h = 1\mathrm{Q}$



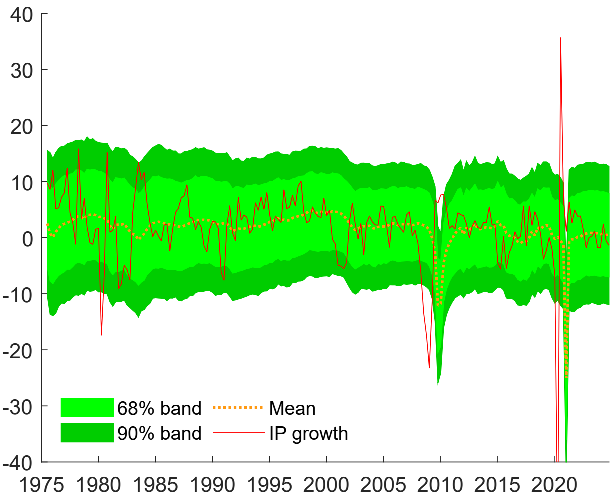
(b) **UCSV** model, $h = 1\mathrm{Q}$



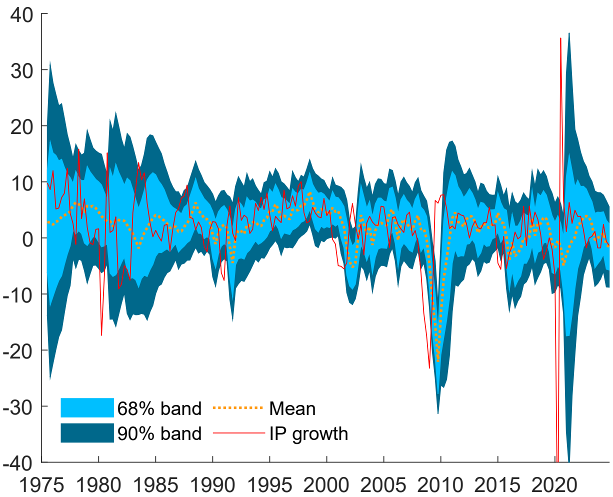
(c) **UC** model, $h = 2\mathrm{Q}$



(d) **UCSV** model, $h = 2\mathrm{Q}$

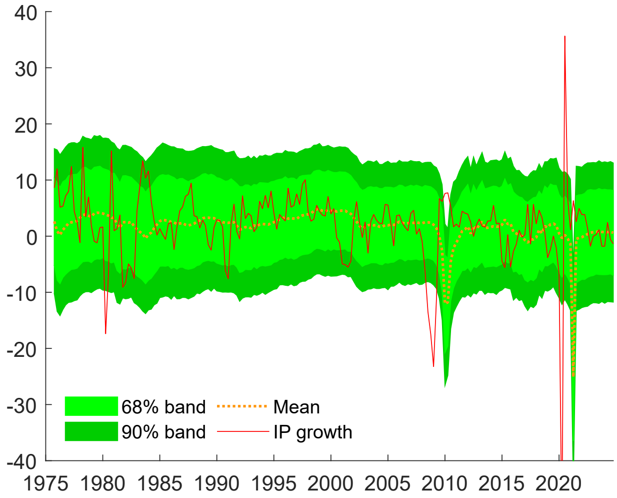


(e) **UC** model, $h = 3\mathrm{Q}$

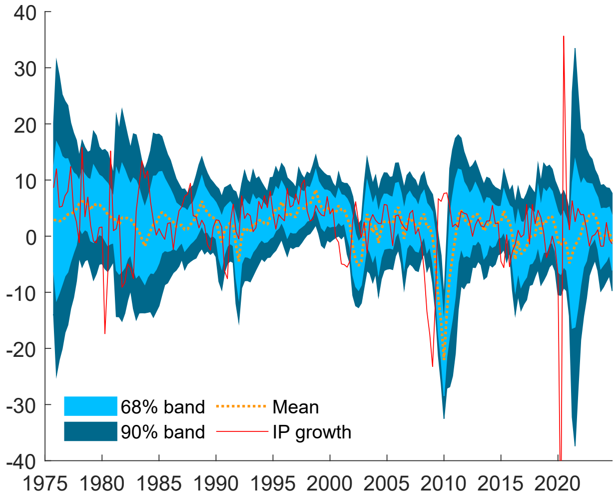


(f) **UCSV** model, $h = 3\mathrm{Q}$

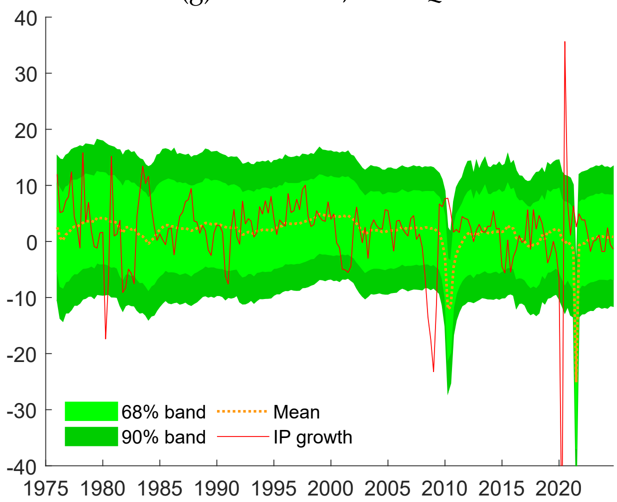
Figure E.3 (cont.): Equal-tailed 68% and 90% predictive bands



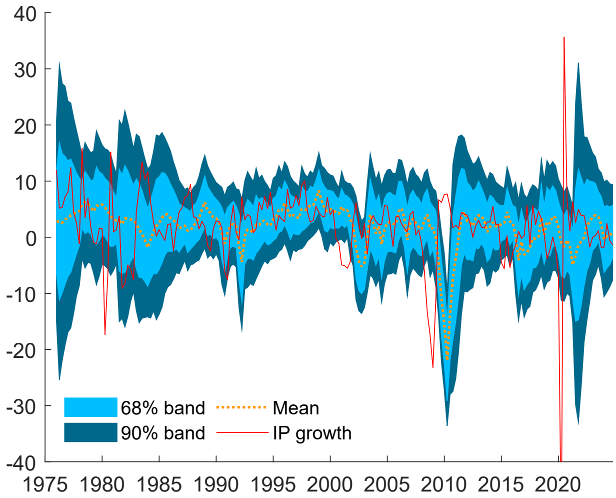
(g) **UC** model, $h = 4Q$



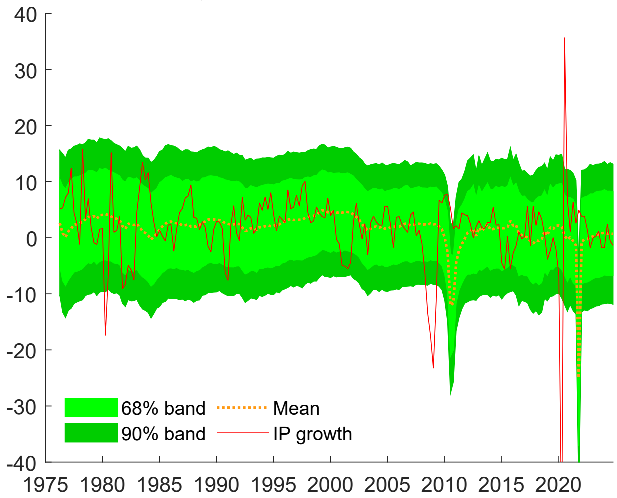
(h) **UCSV** model, $h = 4Q$



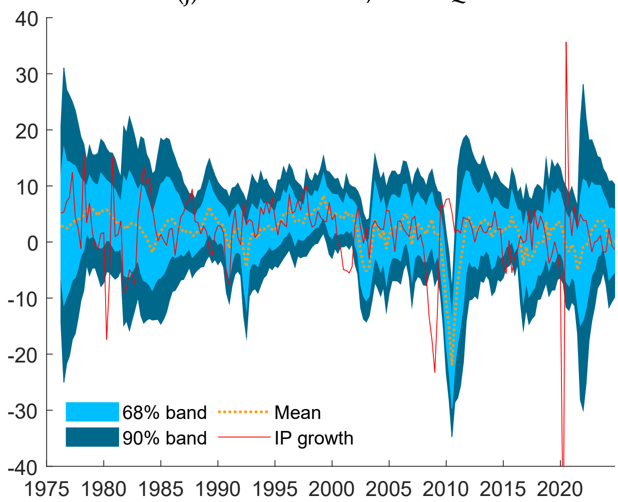
(i) **UC** model, $h = 5Q$



(j) **UCSV** model, $h = 5Q$

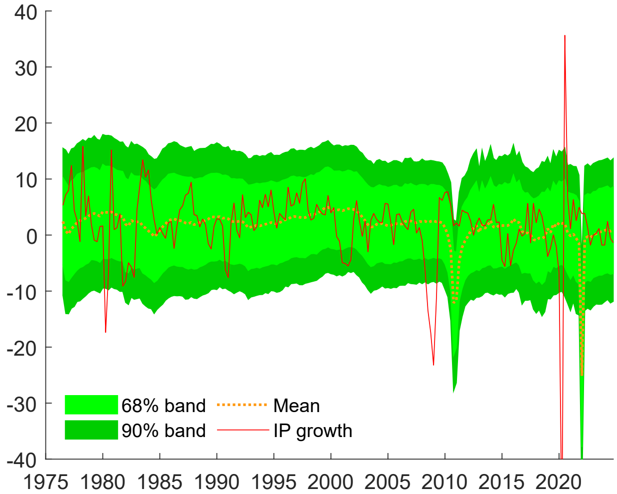


(k) **UC** model, $h = 6Q$

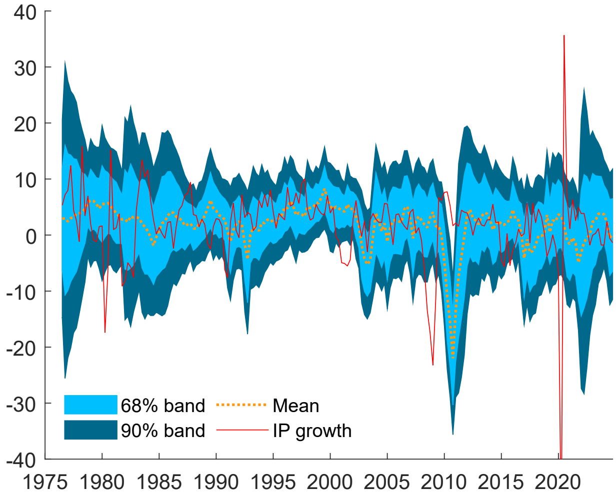


(l) **UCSV** model, $h = 6Q$

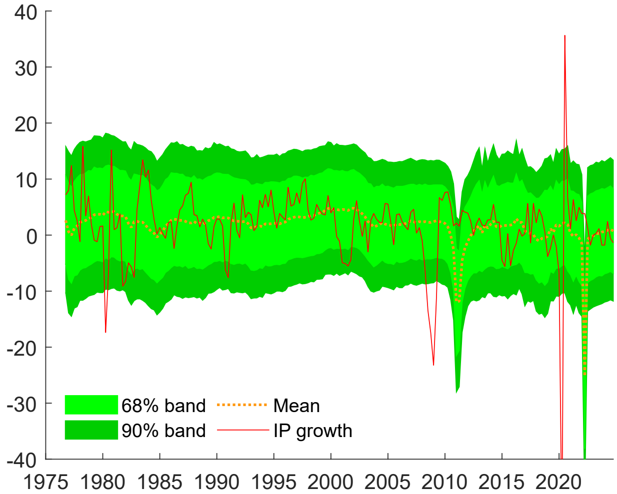
Figure E.3 (cont.): Equal-tailed 68% and 90% predictive bands



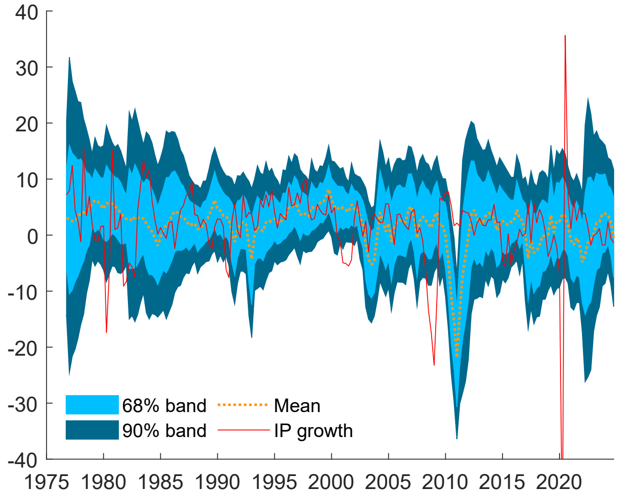
(m) **UC** model, $h = 7Q$



(n) **UCSV** model, $h = 7Q$



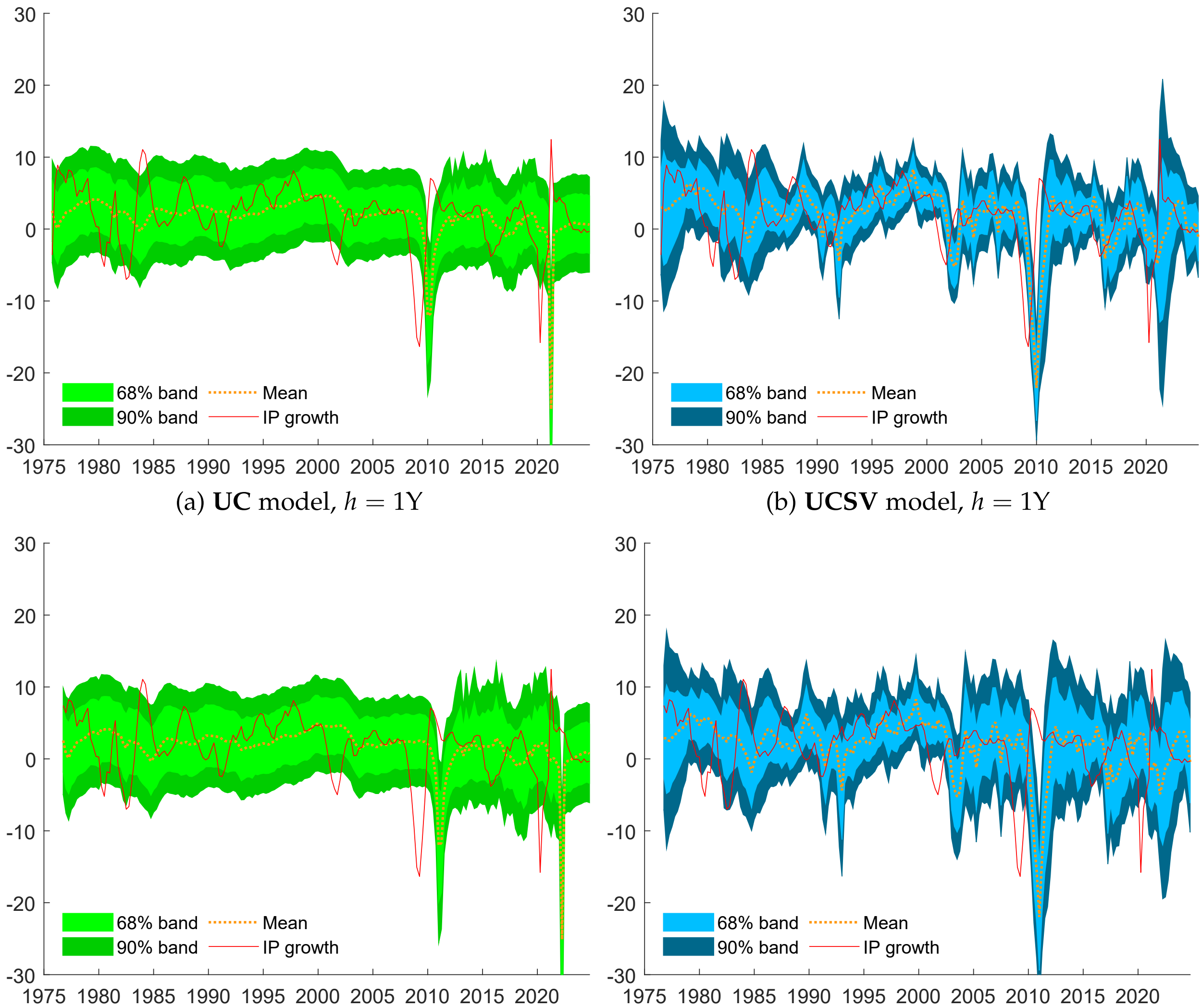
(o) **UC** model, $h = 8Q$



(p) **UCSV** model, $h = 8Q$

Note: The panels show the 68% and 90% predictive bands (with light and dark shades, respectively) along with the mean predictions (dotted lines) and the realized annualized quarter-on-quarter industrial production growth (solid lines) between 1974:Q4 + h and 2024:Q4. Dates on the horizontal axes correspond to the forecast target dates.

Figure E.4: Equal-tailed 68% and 90% predictive bands



(a) **UC** model, $h = 1\text{Y}$

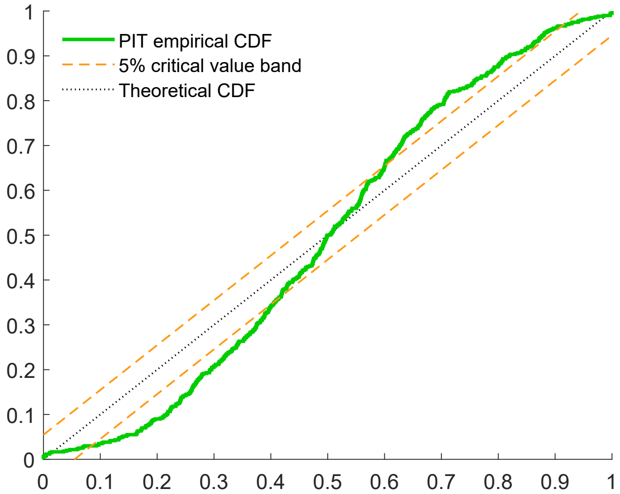
(b) **UCSV** model, $h = 1\text{Y}$

(c) **UC** model, $h = 2\text{Y}$

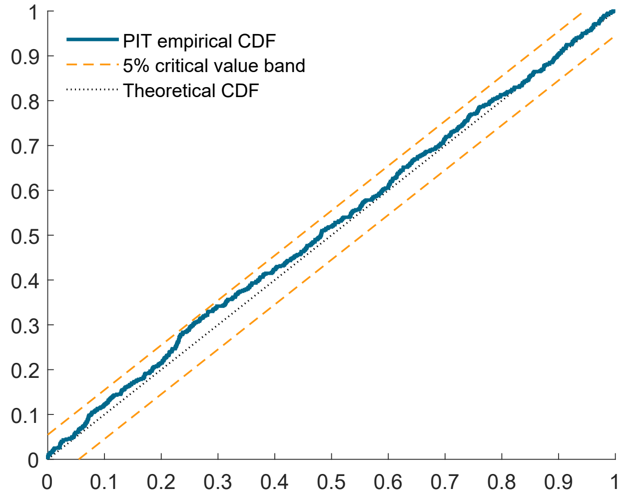
(d) **UCSV** model, $h = 2\text{Y}$

Note: The panels show the 68% and 90% predictive bands (with light and dark shades, respectively) along with the mean predictions (dotted lines) and the realized year-on-year industrial production growth (solid lines) between 1974:Q4 + h and 2024:Q4. Dates on the horizontal axes correspond to the forecast target dates.

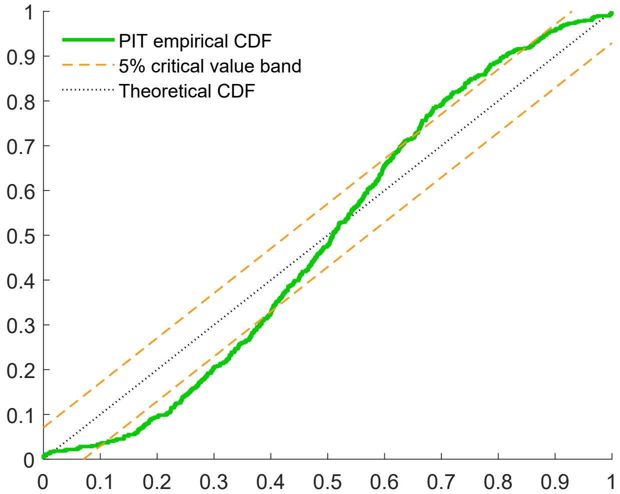
Figure E.5: Empirical CDFs and 5% critical value bands of PITs



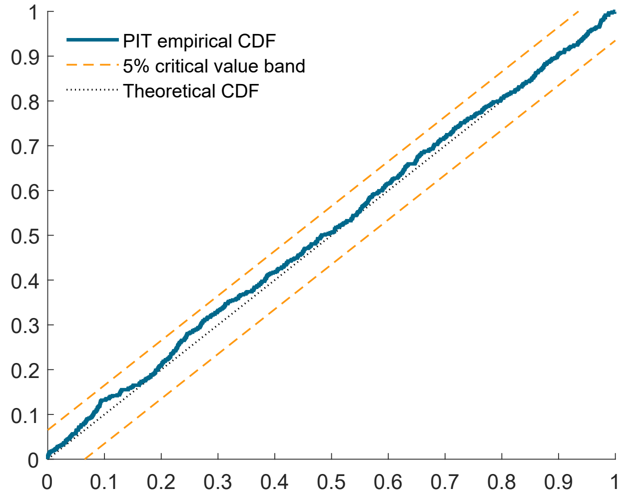
(a) **UC** model, $h = 1M$



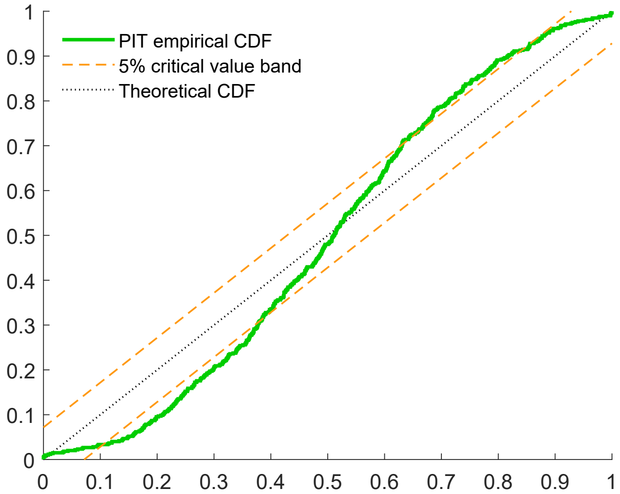
(b) **UCSV** model, $h = 1M$



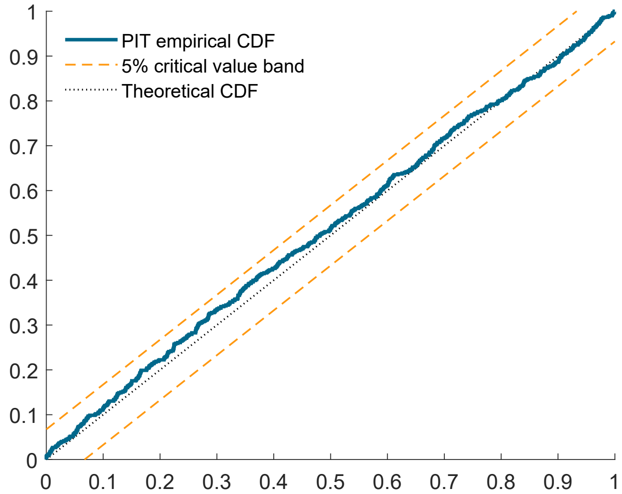
(c) **UC** model, $h = 2M$



(d) **UCSV** model, $h = 2M$

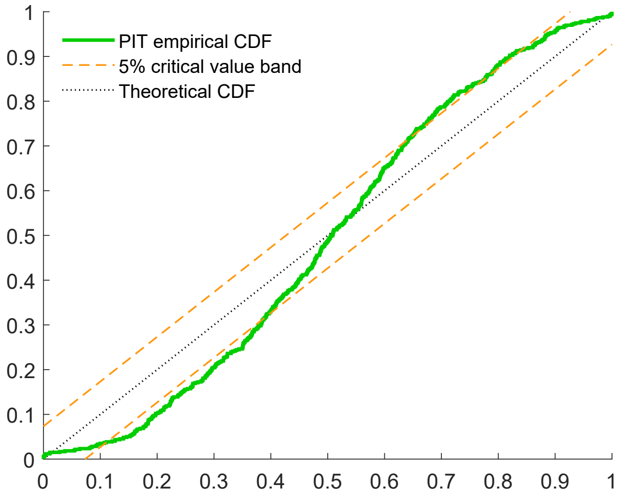


(e) **UC** model, $h = 3M$

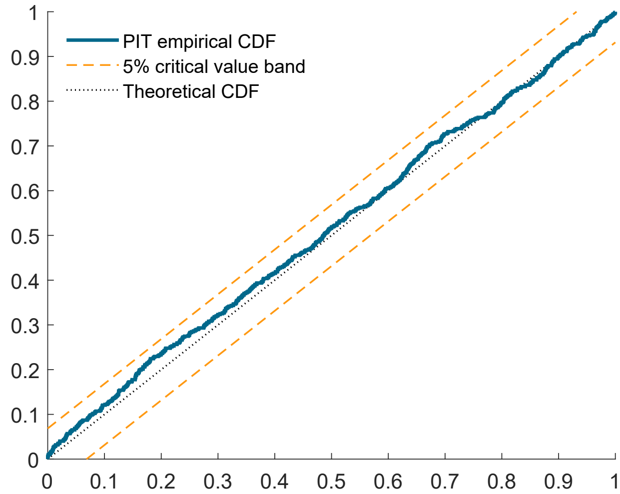


(f) **UCSV** model, $h = 3M$

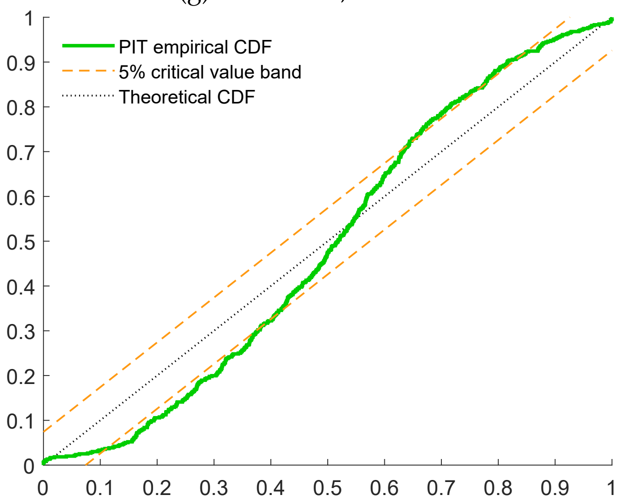
Figure E.5 (cont.): Empirical CDFs and 5% critical value bands of PITs



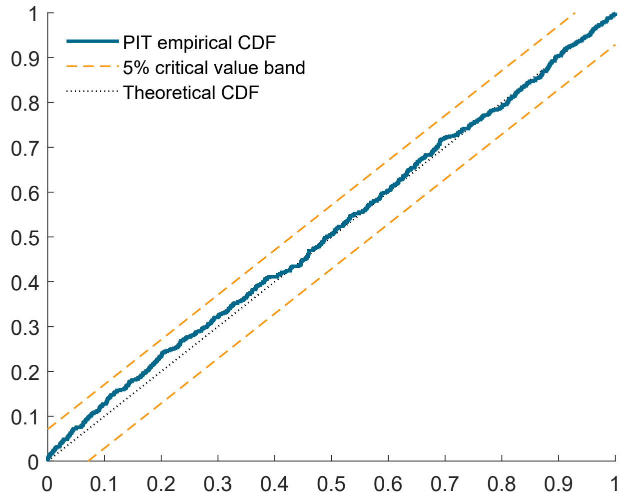
(g) **UC** model, $h = 4M$



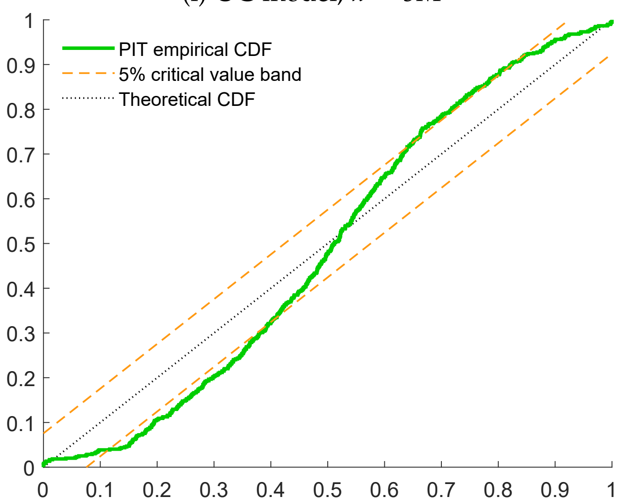
(h) **UCSV** model, $h = 4M$



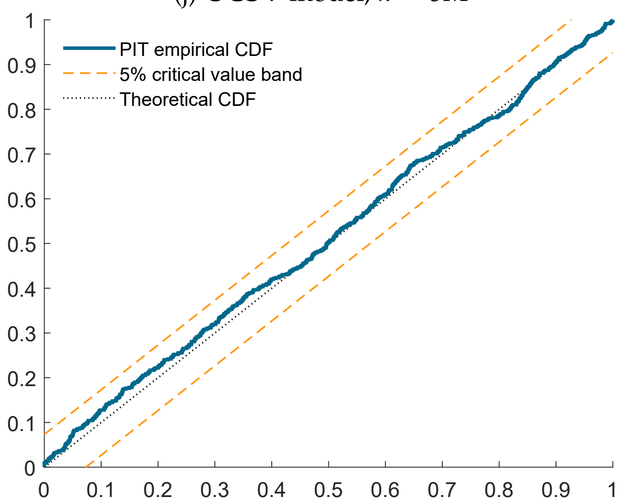
(i) **UC** model, $h = 5M$



(j) **UCSV** model, $h = 5M$

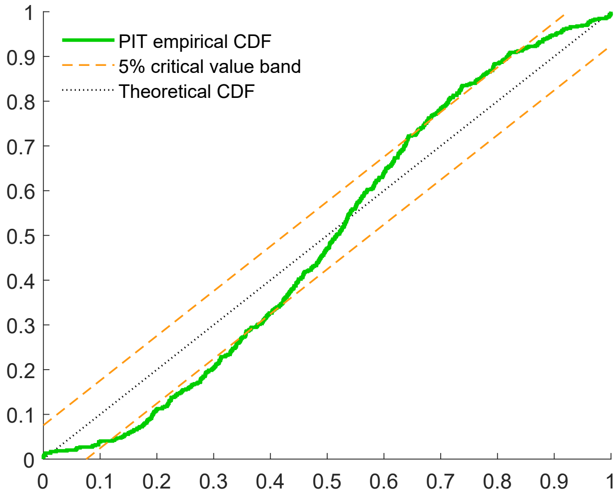


(k) **UC** model, $h = 6M$

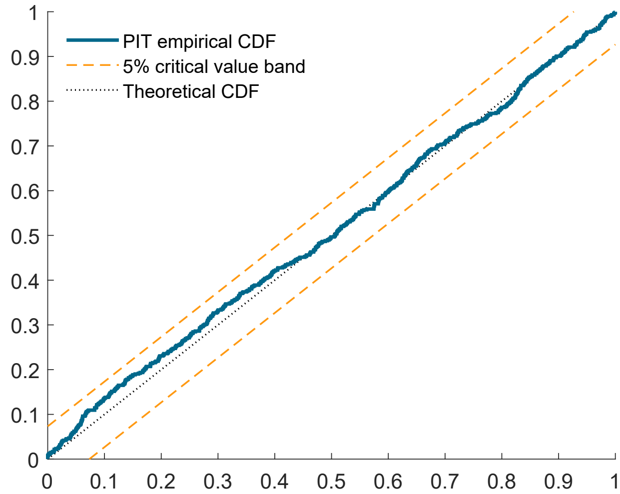


(l) **UCSV** model, $h = 6M$

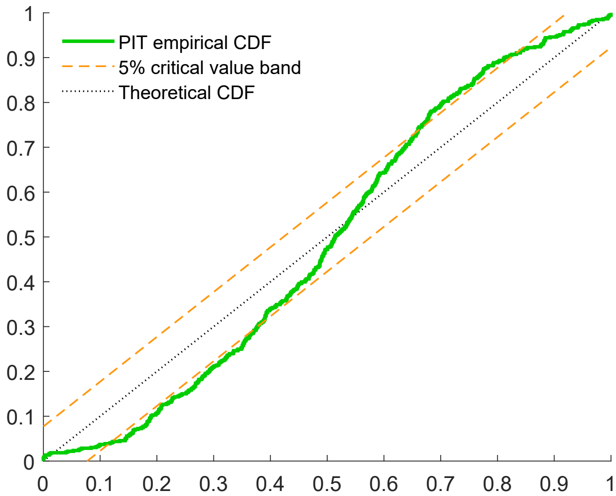
Figure E.5 (cont.): Empirical CDFs and 5% critical value bands of PITs



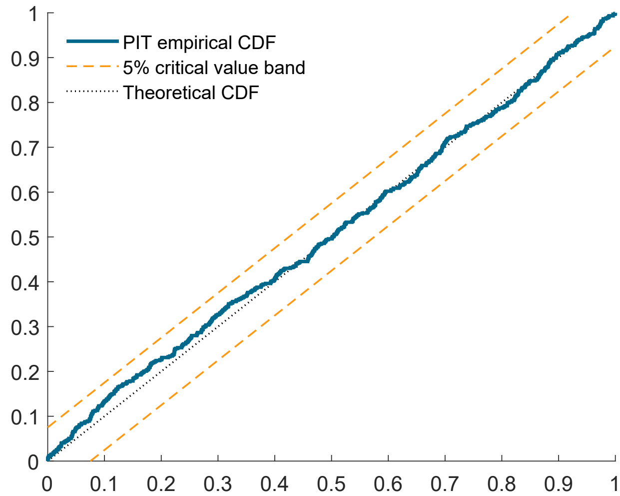
(m) **UC** model, $h = 7\text{M}$



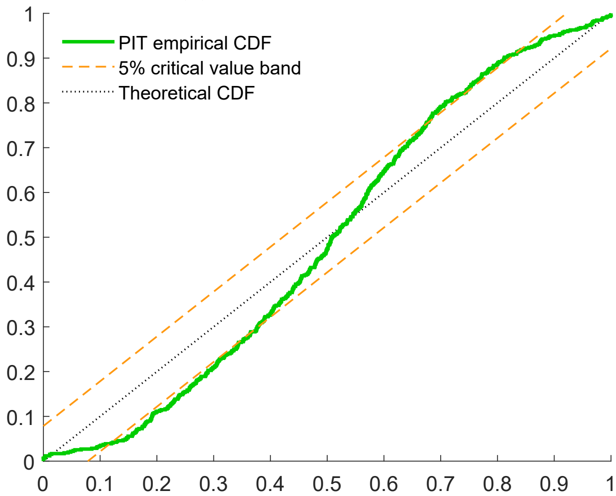
(n) **UCSV** model, $h = 7\text{M}$



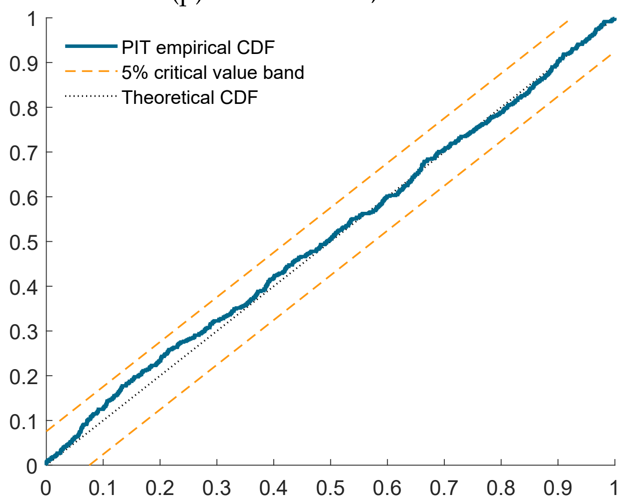
(o) **UC** model, $h = 8\text{M}$



(p) **UCSV** model, $h = 8\text{M}$

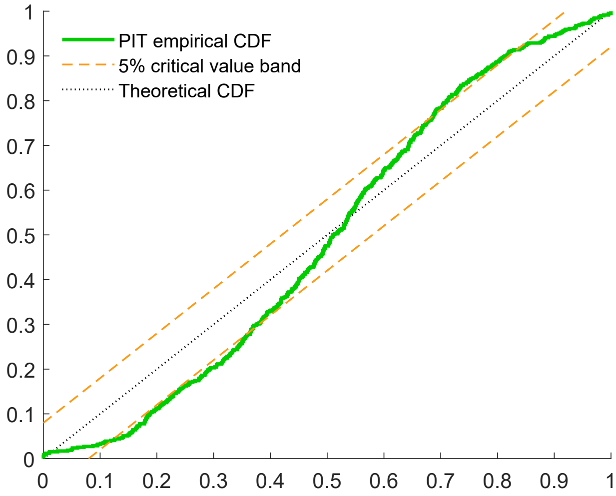


(q) **UC** model, $h = 9\text{M}$

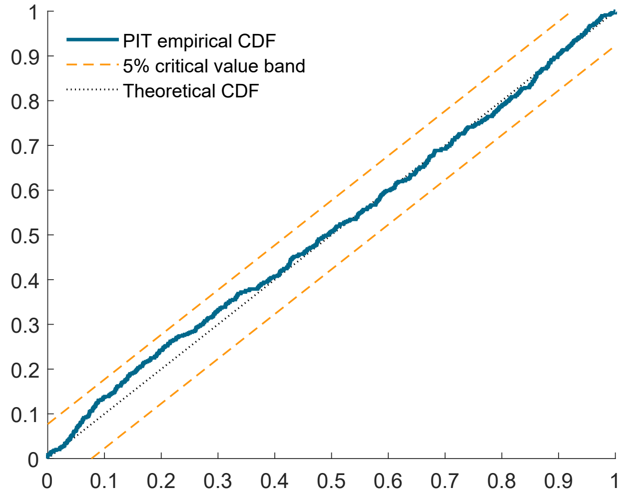


(r) **UCSV** model, $h = 9\text{M}$

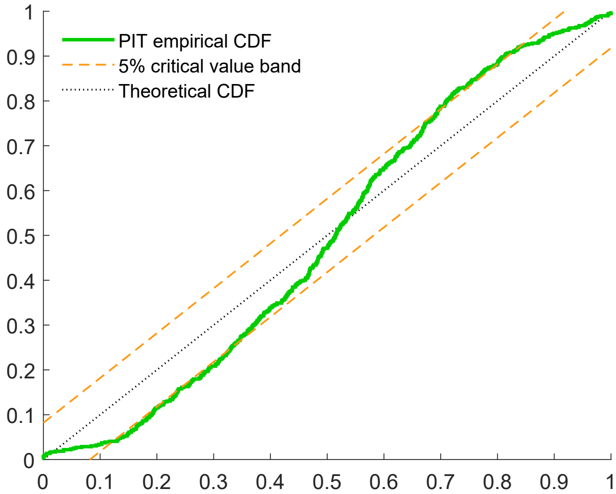
Figure E.5 (cont.): Empirical CDFs and 5% critical value bands of PITs



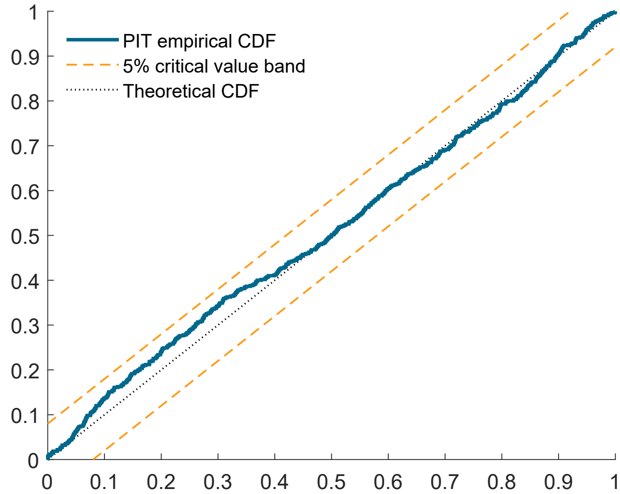
(s) **UC** model, $h = 10$M



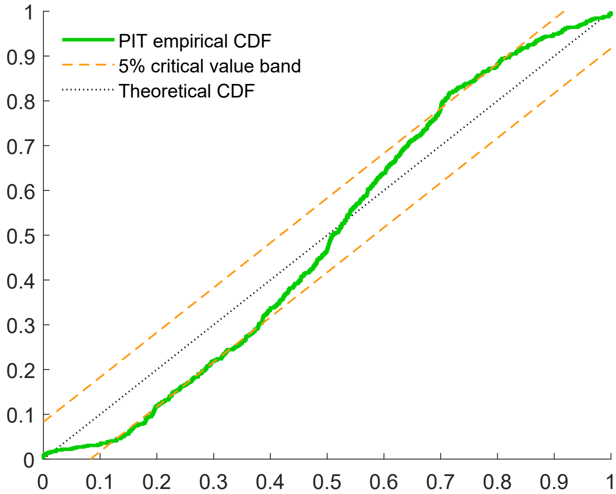
(t) **UCSV** model, $h = 10$M



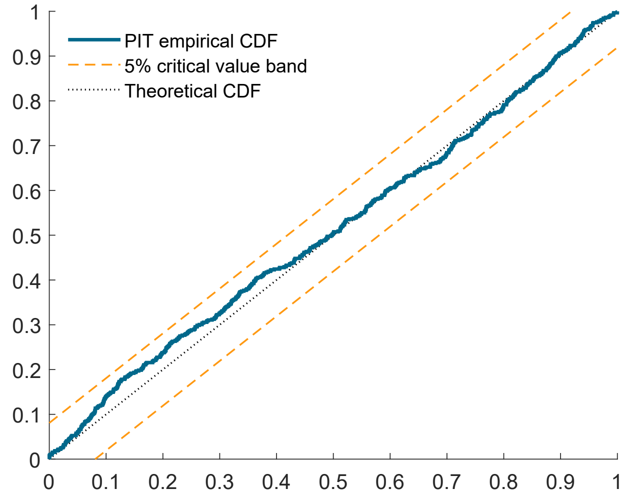
(u) **UC** model, $h = 11$M



(v) **UCSV** model, $h = 11$M



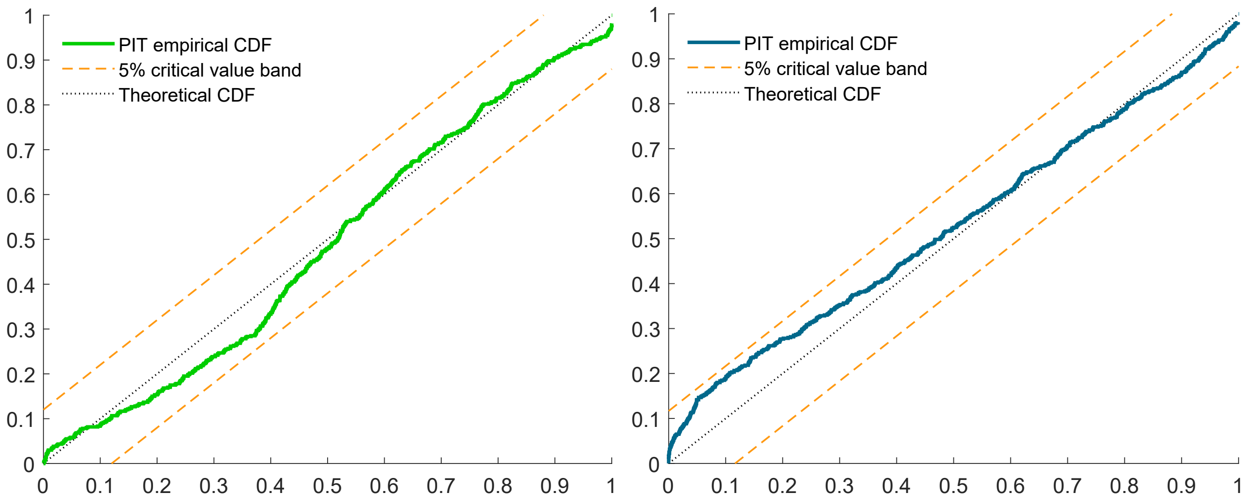
(w) **UC** model, $h = 12$M



(x) **UCSV** model, $h = 12$M

Note: The figures show the empirical CDFs of the PITs (solid lines), the 5% critical bands based on the Kolmogorov–Smirnov-type test (dashed lines), and the 45 degree lines (dotted) corresponding to the theoretical CDF under correct calibration.

Figure E.6: Empirical CDFs and 5% critical value bands of PITs

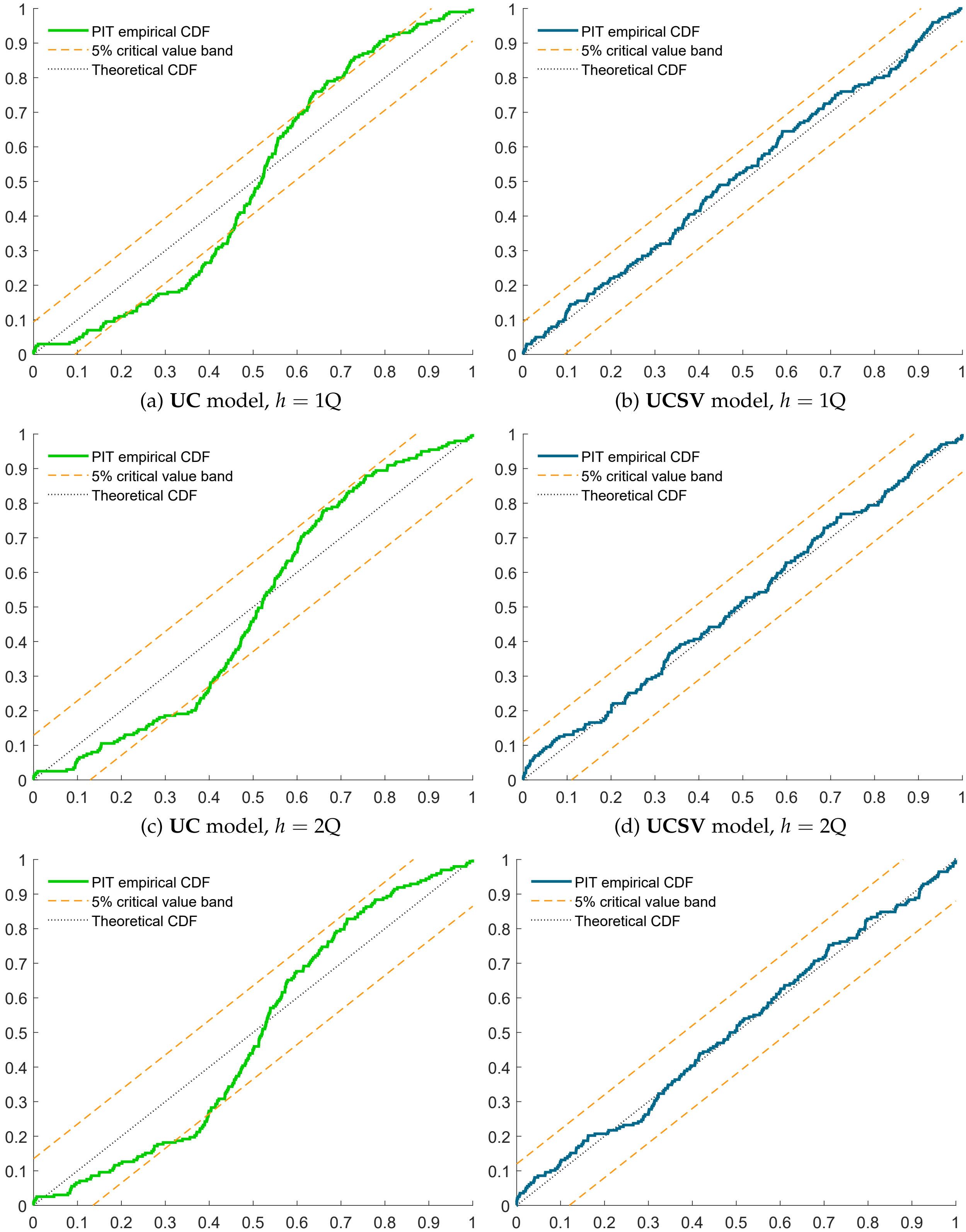


(a) **UC** model, $h = 1\text{Y}$

(b) **UCSV** model, $h = 1\text{Y}$

Note: The figures show the empirical CDFs of the PITs (solid lines), the 5% critical bands based on the Kolmogorov–Smirnov-type test (dashed lines), and the 45 degree lines (dotted) corresponding to the theoretical CDF under correct calibration.

Figure E.7: Empirical CDFs and 5% critical value bands of PITs



(a) **UC** model, $h = 1Q$

(b) **UCSV** model, $h = 1Q$

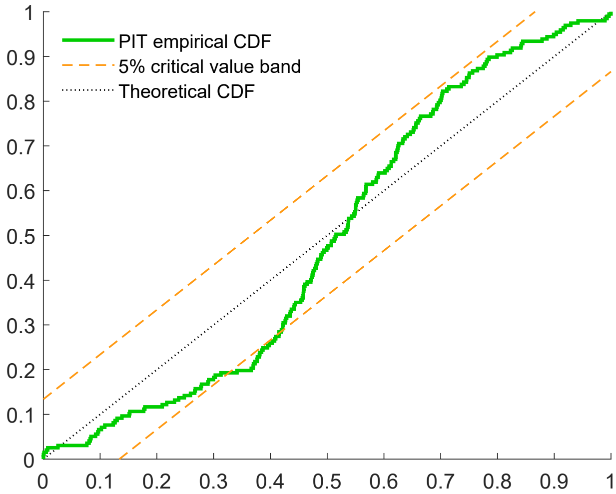
(c) **UC** model, $h = 2Q$

(d) **UCSV** model, $h = 2Q$

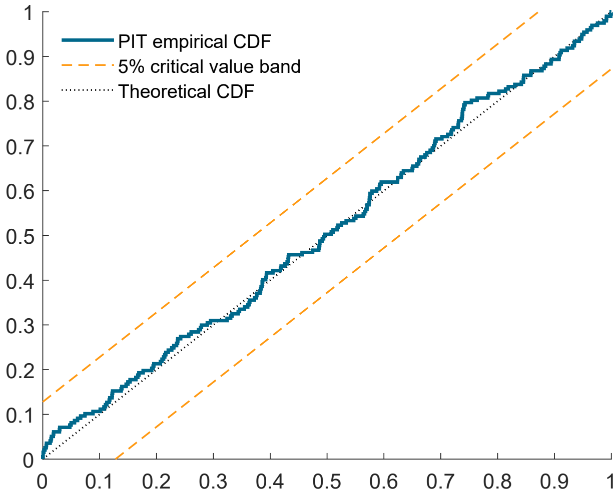
(e) **UC** model, $h = 3Q$

(f) **UCSV** model, $h = 3Q$

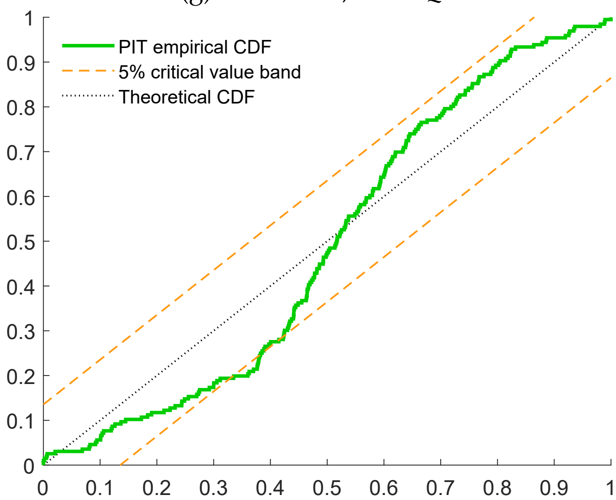
Figure E.7 (cont.): Empirical CDFs and 5% critical value bands of PITs



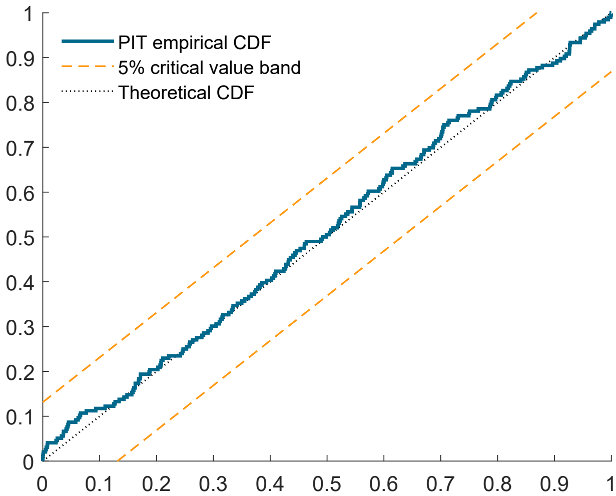
(g) **UC** model, $h = 4Q$



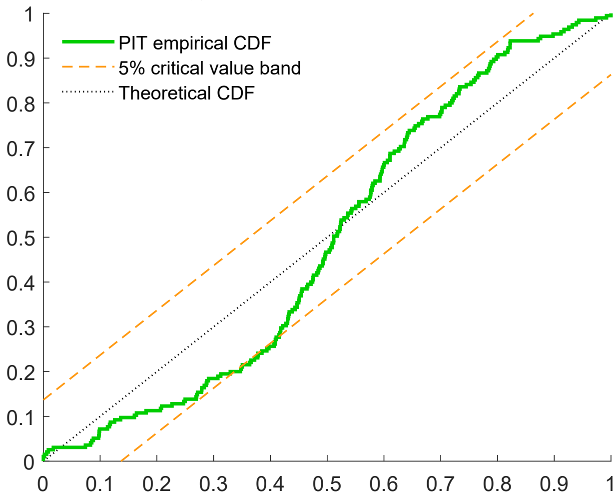
(h) **UCSV** model, $h = 4Q$



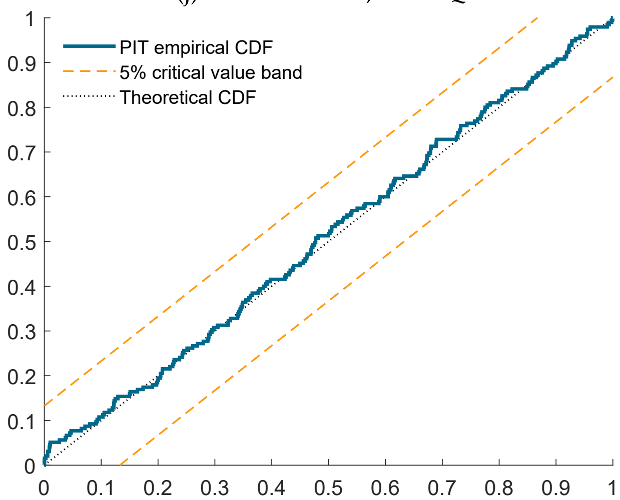
(i) **UC** model, $h = 5Q$



(j) **UCSV** model, $h = 5Q$

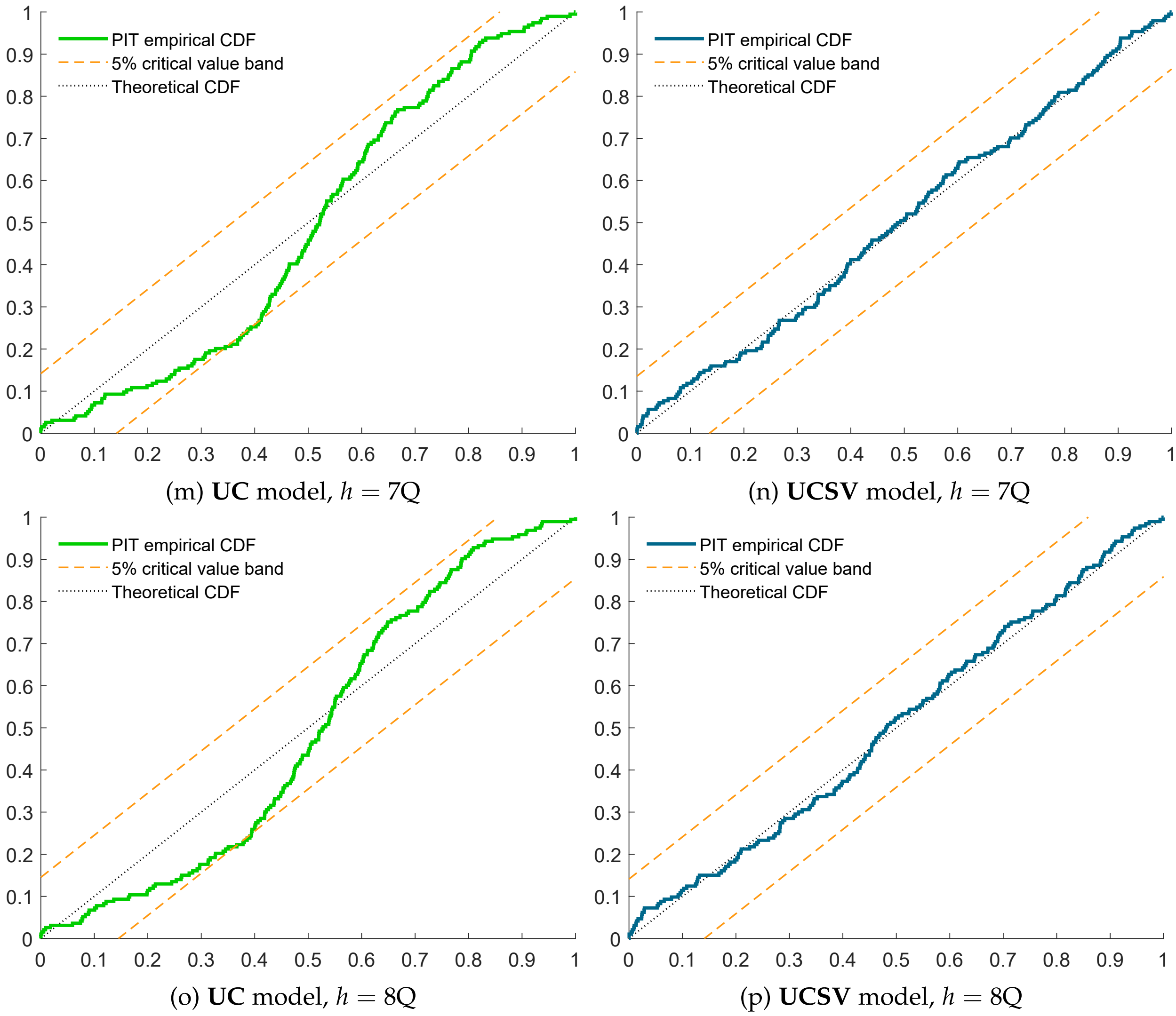


(k) **UC** model, $h = 6Q$



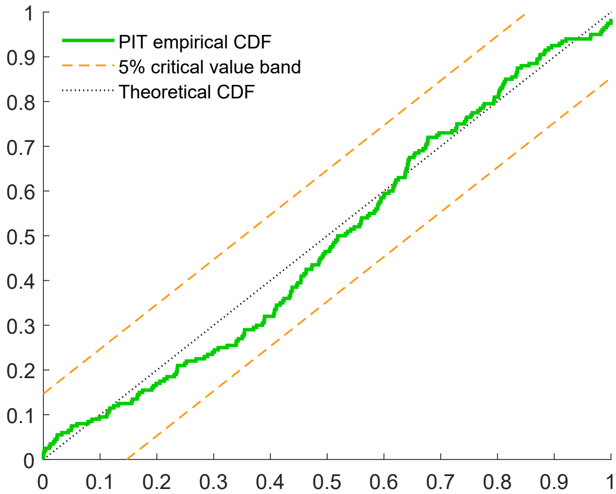
(l) **UCSV** model, $h = 6Q$

Figure E.7 (cont.): Empirical CDFs and 5% critical value bands of PITs

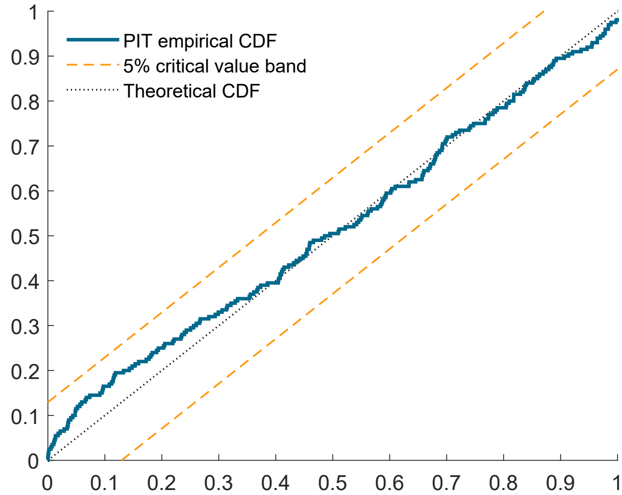


Note: The figures show the empirical CDFs of the PITs (solid lines), the 5% critical bands based on the Kolmogorov–Smirnov-type test (dashed lines), and the 45 degree lines (dotted) corresponding to the theoretical CDF under correct calibration.

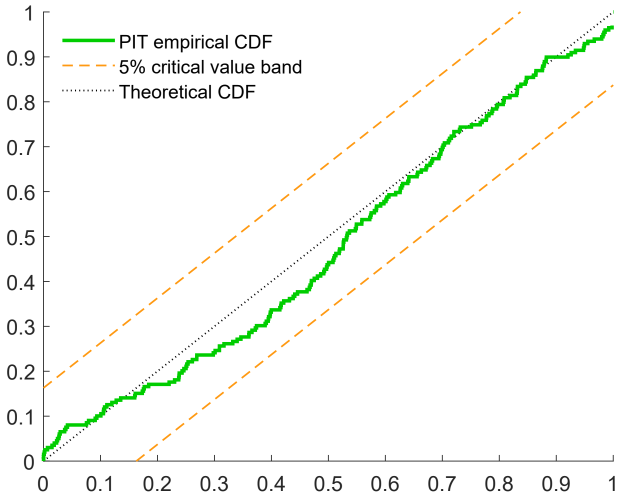
Figure E.8: Empirical CDFs and 5% critical value bands of PITs



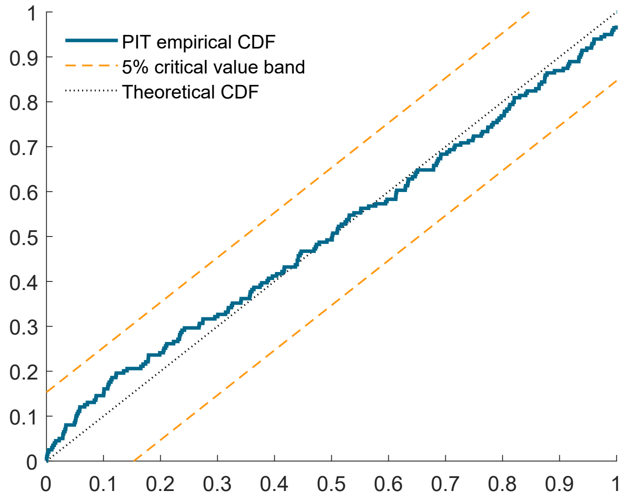
(a) **UC** model, $h = 1Y$



(b) **UCSV** model, $h = 1Y$



(c) **UC** model, $h = 2Y$



(d) **UCSV** model, $h = 2Y$

Note: The figures show the empirical CDFs of the PITs (solid lines), the 5% critical bands based on the Kolmogorov–Smirnov-type test (dashed lines), and the 45 degree lines (dotted) corresponding to the theoretical CDF under correct calibration.

Table E.1: Correct specification tests for monthly US industrial production growth

UC

Region or weight function		KS 1M	KS 2M	KS 3M	KS 4M	KS 5M	KS 6M	KS 7M	KS 8M	KS 9M	KS 10M	KS 11M	KS 12M	KS 1Y	CvM 1M	CvM 2M	CvM 3M	CvM 4M	CvM 5M	CvM 6M	CvM 7M	CvM 8M	CvM 9M	CvM 10M	CvM 11M	CvM 12M	CvM 1Y
Full	$r \in [0, 1]$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.01	**0.24**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.46**
Left tail	$r \in [0, 0.1]$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.57**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.62**
Right tail	$r \in [0.9, 1]$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.00	0.01	**0.21**	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.03	0.04	0.05	0.04	0.03	**0.36**
Left half	$r \in [0, 0.5]$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	**0.17**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.24**
Right half	$r \in [0.5, 1]$	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.00	0.01	0.00	**0.78**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	**0.85**
Center	$r \in [0.1, 0.9]$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.01	**0.24**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.45**
Tails	$r \in \{[0, 0.1] \cup [0.9, 1]\}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.37**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.57**
Left tail	$w(r) = (1-r)^2$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.29**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.31**
Right tail	$w(r) = r^2$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.29**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.66**
Center	$w(r) = r(1-r)$	0.00	0.00	0.01	0.00	0.01	0.01	0.02	0.01	0.03	0.02	0.04	0.04	**0.23**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.44**
Tails	$w(r) = (2r-1)^2$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.20**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.50**

UCSV

Region or weight function		KS 1M	KS 2M	KS 3M	KS 4M	KS 5M	KS 6M	KS 7M	KS 8M	KS 9M	KS 10M	KS 11M	KS 12M	KS 1Y	CvM 1M	CvM 2M	CvM 3M	CvM 4M	CvM 5M	CvM 6M	CvM 7M	CvM 8M	CvM 9M	CvM 10M	CvM 11M	CvM 12M	CvM 1Y
Full	$r \in [0, 1]$	**0.13**	**0.49**	**0.48**	**0.35**	**0.49**	**0.65**	**0.52**	**0.49**	**0.45**	**0.41**	**0.36**	**0.35**	**0.14**	**0.18**	**0.47**	**0.42**	**0.45**	**0.58**	**0.59**	**0.54**	**0.64**	**0.57**	**0.50**	**0.44**	**0.47**	**0.28**
Left tail	$r \in [0, 0.1]$	**0.06**	0.04	**0.25**	**0.25**	**0.18**	**0.22**	**0.10**	**0.18**	**0.17**	**0.10**	**0.15**	**0.14**	0.03	**0.06**	**0.12**	**0.27**	**0.15**	**0.13**	**0.18**	**0.11**	**0.18**	**0.21**	**0.16**	**0.20**	**0.24**	0.02
Right tail	$r \in [0.9, 1]$	**0.78**	**0.80**	**0.54**	**0.62**	**0.94**	**0.85**	**0.91**	**0.81**	**0.88**	**0.65**	**0.56**	**0.41**	**0.32**	**0.64**	**0.81**	**0.64**	**0.84**	**0.97**	**0.89**	**0.95**	**0.83**	**0.90**	**0.65**	**0.51**	**0.47**	**0.25**
Left half	$r \in [0, 0.5]$	**0.08**	**0.38**	**0.38**	**0.27**	**0.39**	**0.52**	**0.40**	**0.38**	**0.35**	**0.32**	**0.29**	**0.28**	**0.12**	**0.08**	**0.33**	**0.28**	**0.31**	**0.38**	**0.42**	**0.33**	**0.40**	**0.36**	**0.31**	**0.27**	**0.29**	**0.16**
Right half	$r \in [0.5, 1]$	**0.65**	**0.68**	**0.76**	**0.58**	**0.77**	**0.62**	**0.92**	**0.96**	**0.98**	**0.93**	**0.87**	**0.82**	**0.74**	**0.50**	**0.64**	**0.71**	**0.68**	**0.92**	**0.78**	**0.94**	**0.97**	**0.96**	**0.95**	**0.91**	**0.88**	**0.76**
Center	$r \in [0.1, 0.9]$	**0.13**	**0.52**	**0.48**	**0.35**	**0.49**	**0.64**	**0.51**	**0.49**	**0.45**	**0.41**	**0.36**	**0.35**	**0.15**	**0.18**	**0.48**	**0.42**	**0.46**	**0.60**	**0.60**	**0.57**	**0.66**	**0.57**	**0.52**	**0.45**	**0.47**	**0.37**
Tails	$r \in \{[0, 0.1] \cup [0.9, 1]\}$	**0.13**	0.04	**0.35**	**0.32**	**0.22**	**0.26**	**0.12**	**0.20**	**0.21**	**0.12**	**0.16**	**0.15**	0.03	**0.16**	**0.19**	**0.40**	**0.25**	**0.22**	**0.30**	**0.18**	**0.28**	**0.35**	**0.23**	**0.27**	**0.31**	0.02
Left tail	$w(r) = (1-r)^2$	**0.08**	**0.08**	**0.33**	**0.16**	**0.21**	**0.29**	**0.15**	**0.17**	**0.17**	**0.17**	**0.18**	**0.13**	0.02	**0.09**	**0.33**	**0.31**	**0.29**	**0.34**	**0.41**	**0.31**	**0.38**	**0.34**	**0.29**	**0.27**	**0.28**	**0.13**
Right tail	$w(r) = r^2$	**0.61**	**0.81**	**0.78**	**0.74**	**0.89**	**0.79**	**0.85**	**0.91**	**0.93**	**0.81**	**0.74**	**0.72**	**0.47**	**0.41**	**0.63**	**0.62**	**0.69**	**0.90**	**0.80**	**0.88**	**0.94**	**0.90**	**0.86**	**0.75**	**0.76**	**0.63**
Center	$w(r) = r(1-r)$	**0.23**	**0.59**	**0.47**	**0.71**	**0.74**	**0.70**	**0.67**	**0.72**	**0.68**	**0.63**	**0.38**	**0.51**	**0.49**	**0.19**	**0.49**	**0.42**	**0.49**	**0.64**	**0.62**	**0.61**	**0.71**	**0.63**	**0.57**	**0.49**	**0.52**	**0.40**
Tails	$w(r) = (2r-1)^2$	**0.16**	**0.07**	**0.42**	**0.33**	**0.27**	**0.22**	**0.11**	**0.24**	**0.26**	**0.16**	**0.18**	**0.14**	0.02	**0.14**	**0.37**	**0.41**	**0.30**	**0.35**	**0.43**	**0.32**	**0.40**	**0.36**	**0.30**	**0.28**	**0.30**	**0.09**

Note: The table reports the p-values of the correct specification of predictive densities of the UC and UCSV models based on the Kolmogorov–Smirnov- and Cramér–von Mises-type (KS and CvM, respectively) test statistics in various regions or with various weight functions (in rows) at forecast horizons of $h = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$ months (annualized MoM growth) and $h = 1$ year (YoY growth) in columns. Boldface indicates entries for which the null hypothesis of correct calibration is not rejected at the $\alpha = 5\%$ significance level. The forecast evaluation period spans from 1974:M12 + h to 2024:M12. For $h = 1$, the p-values are computed using the limiting Gaussian process with 10,000 simulations. For $h > 1$, the critical values are based on the bootstrap with block length $\ell = \max\{h-1, \lfloor P^{1/3} \rfloor\}$. The number of bootstrap replications is 10,000, the domain of r is discretized over $[0 : 0.001 : 1]$.

Table E.2: Correct specification tests for quarterly US industrial production growth

UC

Region or weight function		KS 1Q	KS 2Q	KS 3Q	KS 4Q	KS 5Q	KS 6Q	KS 7Q	KS 8Q	KS 1Y	KS 2Y	CvM 1Q	CvM 2Q	CvM 3Q	CvM 4Q	CvM 5Q	CvM 6Q	CvM 7Q	CvM 8Q	CvM 1Y	CvM 2Y
Full	$r \in [0, 1]$	0.00	0.01	0.01	0.01	0.01	0.03	0.03	0.03	**0.44**	**0.50**	0.00	0.00	0.01	0.01	0.01	0.01	0.02	0.02	**0.53**	**0.58**
Left tail	$r \in [0, 0.1]$	0.01	0.01	**0.10**	**0.12**	**0.10**	**0.12**	**0.17**	**0.24**	**0.47**	**0.43**	0.02	0.03	**0.13**	**0.15**	**0.14**	**0.14**	**0.21**	**0.29**	**0.54**	**0.51**
Right tail	$r \in [0.9, 1]$	0.01	0.02	0.03	0.03	0.01	0.01	0.00	0.00	**0.34**	**0.33**	0.01	**0.09**	**0.06**	0.02	0.03	0.01	0.00	0.00	**0.41**	**0.38**
Left half	$r \in [0, 0.5]$	0.00	0.00	0.00	0.00	0.01	0.02	0.02	0.02	**0.31**	**0.36**	0.00	0.01	0.01	0.01	0.01	0.01	0.02	0.02	**0.32**	**0.35**
Right half	$r \in [0.5, 1]$	0.00	0.05	**0.10**	**0.07**	**0.14**	**0.09**	**0.14**	**0.10**	**0.87**	**0.62**	0.00	0.02	0.05	0.04	**0.07**	**0.05**	**0.08**	0.05	**0.79**	**0.84**
Center	$r \in [0.1, 0.9]$	0.00	0.01	0.01	0.01	0.01	0.03	0.03	0.03	**0.44**	**0.50**	0.00	0.00	0.01	0.01	0.01	0.01	0.02	0.02	**0.52**	**0.57**
Tails	$r \in \{[0, 0.1] \cup [0.9, 1]\}$	0.02	0.02	**0.12**	**0.10**	**0.06**	**0.08**	0.05	0.04	**0.64**	**0.61**	0.00	0.03	**0.06**	0.04	0.05	0.03	0.02	0.03	**0.60**	**0.54**
Left tail	$w(r) = (1-r)^2$	0.00	0.01	0.01	0.01	0.01	0.01	0.02	0.02	**0.44**	**0.57**	0.00	0.01	0.01	0.01	0.01	0.01	0.02	0.02	**0.42**	**0.47**
Right tail	$w(r) = r^2$	0.00	0.00	0.01	0.01	0.01	0.00	0.01	0.00	**0.64**	**0.52**	0.00	0.01	0.01	0.01	0.02	0.02	0.02	0.01	**0.65**	**0.71**
Center	$w(r) = r(1-r)$	0.00	0.01	0.01	0.01	0.02	0.03	0.04	0.04	**0.39**	**0.44**	0.00	0.01	0.01	0.01	0.02	0.02	0.03	0.02	**0.50**	**0.53**
Tails	$w(r) = (2r-1)^2$	0.01	0.04	**0.07**	0.04	0.03	0.02	0.03	0.02	**0.58**	**0.50**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.64**	**0.76**

UCSV

Region or weight function		KS 1Q	KS 2Q	KS 3Q	KS 4Q	KS 5Q	KS 6Q	KS 7Q	KS 8Q	KS 1Y	KS 2Y	CvM 1Q	CvM 2Q	CvM 3Q	CvM 4Q	CvM 5Q	CvM 6Q	CvM 7Q	CvM 8Q	CvM 1Y	CvM 2Y
Full	$r \in [0, 1]$	**0.58**	**0.83**	**0.91**	**0.76**	**0.90**	**0.94**	**0.92**	**0.90**	**0.43**	**0.61**	**0.58**	**0.80**	**0.88**	**0.93**	**0.94**	**0.97**	**0.93**	**0.87**	**0.60**	**0.71**
Left tail	$r \in [0, 0.1]$	**0.43**	**0.25**	**0.34**	**0.30**	**0.35**	**0.33**	**0.42**	**0.29**	**0.12**	**0.20**	**0.41**	**0.16**	**0.34**	**0.34**	**0.34**	**0.45**	**0.41**	**0.39**	**0.09**	**0.20**
Right tail	$r \in [0.9, 1]$	**0.39**	**0.66**	**0.56**	**0.92**	**0.78**	**0.72**	**0.34**	**0.35**	**0.36**	**0.35**	**0.37**	**0.52**	**0.72**	**0.93**	**0.82**	**0.70**	**0.46**	**0.27**	**0.36**	**0.27**
Left half	$r \in [0, 0.5]$	**0.66**	**0.70**	**0.79**	**0.83**	**0.86**	**0.86**	**0.89**	**0.82**	**0.34**	**0.50**	**0.61**	**0.71**	**0.76**	**0.86**	**0.93**	**0.96**	**0.86**	**0.80**	**0.38**	**0.57**
Right half	$r \in [0.5, 1]$	**0.42**	**0.65**	**0.76**	**0.58**	**0.75**	**0.82**	**0.78**	**0.83**	**0.90**	**0.82**	**0.45**	**0.74**	**0.85**	**0.87**	**0.82**	**0.88**	**0.86**	**0.79**	**0.93**	**0.76**
Center	$r \in [0.1, 0.9]$	**0.58**	**0.81**	**0.89**	**0.75**	**0.89**	**0.94**	**0.91**	**0.92**	**0.42**	**0.60**	**0.58**	**0.89**	**0.89**	**0.94**	**0.96**	**0.98**	**0.94**	**0.90**	**0.70**	**0.77**
Tails	$r \in \{[0, 0.1] \cup [0.9, 1]\}$	**0.62**	**0.31**	**0.50**	**0.47**	**0.52**	**0.48**	**0.58**	**0.41**	**0.13**	**0.31**	**0.47**	**0.19**	**0.52**	**0.56**	**0.53**	**0.62**	**0.52**	**0.40**	**0.12**	**0.23**
Left tail	$w(r) = (1-r)^2$	**0.42**	**0.30**	**0.43**	**0.42**	**0.49**	**0.42**	**0.51**	**0.38**	**0.14**	**0.29**	**0.61**	**0.65**	**0.75**	**0.85**	**0.90**	**0.95**	**0.86**	**0.84**	**0.34**	**0.54**
Right tail	$w(r) = r^2$	**0.68**	**0.69**	**0.79**	**0.56**	**0.79**	**0.90**	**0.64**	**0.66**	**0.56**	**0.46**	**0.53**	**0.81**	**0.90**	**0.92**	**0.90**	**0.95**	**0.93**	**0.82**	**0.90**	**0.75**
Center	$w(r) = r(1-r)$	**0.45**	**0.82**	**0.90**	**0.82**	**0.90**	**0.93**	**0.85**	**0.86**	**0.86**	**0.87**	**0.55**	**0.85**	**0.90**	**0.95**	**0.95**	**0.97**	**0.93**	**0.88**	**0.73**	**0.80**
Tails	$w(r) = (2r-1)^2$	**0.57**	**0.23**	**0.41**	**0.33**	**0.47**	**0.29**	**0.44**	**0.30**	**0.13**	**0.33**	**0.66**	**0.60**	**0.75**	**0.82**	**0.87**	**0.93**	**0.87**	**0.80**	**0.30**	**0.46**

Note: The table reports the p-values of the correct specification of predictive densities of the UC and UCSV models based on the Kolmogorov–Smirnov- and Cramér–von Mises-type (KS and CvM, respectively) test statistics in various regions or with various weight functions (in rows) at forecast horizons of $h = 1, 2, 3, 4, 5, 6, 7, 8$ quarters (annualized QoQ growth) and $h = 1, 1$ years (YoY growth) in columns. Boldface indicates entries for which the null hypothesis of correct calibration is not rejected at the $\alpha = 5\%$ significance level. The forecast evaluation period spans from 1974:Q4 + h to 2024:Q4. For $h = 1$, the p-values are computed using the finite-sample standard uniform distribution with 10,000 simulations. For $h > 1$, the critical values are based on the bootstrap with block length $\ell = \max\{h-1, \lfloor P^{1/3} \rfloor\}$. The number of bootstrap replications is 10,000, the domain of r is discretized over $[0 : 0.001 : 1]$.

BANCO DE ESPAÑA PUBLICATIONS

WORKING PAPERS

2420 FRANCISCO GONZÁLEZ, JOSÉ E. GUTIÉRREZ and JOSÉ MARÍA SERENA: Shadow seniority? Lending relationships and borrowers' selective default.

2421 ROBERTO BLANCO, MIGUEL GARCÍA-POSADA, SERGIO MAYORDOMO and MARÍA RODRÍGUEZ-MORENO: Access to credit and firm survival during a crisis: the case of zero-bank-debt firms.

2422 FERNANDO CEREZO, PABLO GIRÓN, MARÍA T. GONZÁLEZ-PÉREZ and ROBERTO PASCUAL: The impact of sovereign debt purchase programmes. A case study: the Spanish-to-Portuguese bond yield spread.

2423 EDGAR SILGADO-GÓMEZ: Sovereign uncertainty.

2424 CLODOMIRO FERREIRA, JULIO GÁLVEZ and MYROSLAV PIDKUYKO: Housing tenure, consumption and household debt: life-cycle dynamics during a housing bust in Spain.

2425 RUBÉN DOMÍNGUEZ-DÍAZ and SAMUEL HURTADO: Green energy transition and vulnerability to external shocks.

2426 JOSEP GISBERT and JOSÉ E. GUTIÉRREZ: Bridging the gap? Fintech and financial inclusion.

2427 RODOLFO G. CAMPOS, MARIO LARCH, JACOPO TIMINI, ELENA VIDAL and YOTO V. YOTOV: Does the WTO Promote Trade? A Meta-analysis.

2428 SONER BASKAYA, JOSÉ E. GUTIÉRREZ, JOSÉ MARÍA SERENA and SERAFEIM TSOUKAS: Bank supervision and non-performing loan cleansing.

2429 TODD E. CLARK, GERGELY GANICS, and ELMAR MERTENS: Constructing fan charts from the ragged edge of SPF forecasts.

2430 MIGUEL GARCÍA-POSADA and PETER PAZ: The transmission of monetary policy to credit supply in the euro area.

2431 KLODIANA ISTREFI, FLORENS ODENDAHL and GIULIA SESTIERI: ECB communication and its impact on financial markets.

2432 FRUCTUOSO BORRALLO, LUCÍA CUADRO-SÁEZ, CORINNA GHIRELLI and JAVIER J. PÉREZ: "El Niño" and "La Niña": Revisiting the impact on food commodity prices and euro area consumer prices.

2433 VÍCTOR CABALLERO, CORINNA GHIRELLI, ÁNGEL LUIS GÓMEZ and JAVIER J. PÉREZ: The public-private wage GAP in the euro area a decade after the sovereign debt crisis.

2434 LIDIA CRUCES, ISABEL MICÓ-MILLÁN and SUSANA PÁRRAGA: Female financial portfolio choices and marital property regimes.

2435 RODOLFO G. CAMPOS, ANA-SIMONA MANU, LUIS MOLINA and MARTA SUÁREZ-VARELA: China's financial spillovers to emerging markets.

2436 LUDOVIC PANON, LAURA LEBASTARD, MICHELE MANCINI, ALESSANDRO BORIN, PEONARE CAKA, GIANMARCO CARIOLA, DENNIS ESSERS, ELENA GENTILI, ANDREA LINARELLO, TULLIA PADELLINI, FRANCISCO REQUENA and JACOPO TIMINI: Inputs in Distress: Geoeconomic Fragmentation and Firms' Sourcing.

2437 DANIEL DEJUAN-BITRIA, WAYNE R. LANDSMAN, SERGIO MAYORDOMO and IRENE ROIBÁS: How do changes in financial reporting standards affect relationship lending?

2438 ALICIA AGUILAR and RICARDO GIMENO: Discrete Probability Forecasts: What to expect when you are expecting a monetary policy decision.

2439 RODOLFO G. CAMPOS, JESÚS FERNÁNDEZ-VILLAVERDE, GALO NUÑO and PETER PAZ: Navigating by Falling Stars: Monetary Policy with Fiscally Driven Natural Rates.

2440 ALEJANDRO CASADO and DAVID MARTÍNEZ-MIERA: Local lending specialization and monetary policy.

2441 JORGE ABAD, DAVID MARTÍNEZ-MIERA and JAVIER SUÁREZ: A macroeconomic model of banks' systemic risk taking.

2442 JOSEP PIJOAN-MAS and PAU ROLDAN-BLANCO: Dual labor markets and the equilibrium distribution of firms.

2443 OLYMPIA BOVER, LAURA HOSPIDO and ANA LAMO: Gender and Career Progression: Evidence from the Banco de España.

2444 JESÚS FERNÁNDEZ-VILLAVERDE, GALO NUÑO and JESSE PERLA: Taming the curse of dimensionality: quantitative economics with deep learning.

2445 CLODOMIRO FERREIRA and STEFANO PICA: Households' subjective expectations: disagreement, common drivers and reaction to monetary policy.

2446 ISABEL MICÓ-MILLÁN: Inheritance Tax Avoidance Through the Family Firm.

2447 MIKEL BEDAYO, EVA VALDEOLIVAS and CARLOS PÉREZ: The stabilizing role of local claims in local currency on the variation of foreign claims.

2501 HENRIQUE S. BASSO, MYROSLAV PIDKUYKO and OMAR RACHEDI: Opening the black box: aggregate implications of public investment heterogeneity.

2502 MARCO BARDOSCIA, ADRIAN CARRO, MARC HINTERSCHWEIGER, MAURO NAPOLETANO, LILIT POPOYAN, ANDREA ROVENTINI and ARZU ULUC: The impact of prudential regulations on the UK housing market and economy: insights from an agent-based model.

2503 IRINA BALTEANU, KATJA SCHMIDT and FRANCESCA VIANI: Sourcing all the eggs from one basket: trade dependencies and import prices.

2504 RUBÉN VEIGA DUARTE, SAMUEL HURTADO, PABLO A. AGUILAR GARCÍA, JAVIER QUINTANA GONZÁLEZ and CAROLINA MENÉNDEZ ÁLVAREZ: CATALIST: A new, bigger, better model for evaluating climate change transition risks at Banco de España.

2505 PILAR GARCÍA and DIEGO TORRES: Perceiving central bank communications through press coverage.

2506 MAR DELGADO-TÉLLEZ, JAVIER QUINTANA and DANIEL SANTABÁRBARA: Carbon pricing, border adjustment and renewable energy investment: a network approach.

2507 MARTA GARCÍA RODRÍGUEZ: The role of wage expectations in the labor market.

2508 REBECA ANGUREN, GABRIEL JIMÉNEZ and JOSÉ-LUIS PEYDRÓ: Bank capital requirements and risk-taking: evidence from Basel III.

2509 JORGE E. GALÁN: Macroprudential policy and the tail risk of credit growth.

2510 PETER KARADI, ANTON NAKOV, GALO NUÑO, ERNESTO PASTÉN and DOMINIK THALER: Strike while the Iron is Hot: Optimal Monetary Policy with a Nonlinear Phillips Curve.

2511 MATTEO MOGLIANI and FLORENS ODENDAHL: Density forecast transformations.

2512 LUCÍA LÓPEZ, FLORENS ODENDAHL, SUSANA PÁRRAGA and EDGAR SILGADO-GÓMEZ: The pass-through to inflation of gas price shocks.

2513 CARMEN BROTO and OLIVIER HUBERT: Desertification in Spain: Is there any impact on credit to firms?

2514 ANDRÉS ALONSO-ROBISCO, JOSÉ MANUEL CARBÓ, PEDRO JESÚS CUADROS-SOLAS and JARA QUINTANERO: The effects of open banking on fintech providers: evidence using microdata from Spain.

2515 RODOLFO G. CAMPOS and JACOPO TIMINI: Trade bloc enlargement when many countries join at once.

2516 CORINNA GHIRELLI, JAVIER J. PÉREZ and DANIEL SANTABÁRBARA: Inflation and growth forecast errors and the sacrifice ratio of monetary policy in the euro area.

2517 KOSUKE AOKI, ENRIC MARTORELL and KALIN NIKOLOV: Monetary policy, bank leverage and systemic risk-taking.

2518 RICARDO BARAHONA: Index fund flows and fund distribution channels.

2519 ALVARO FERNÁNDEZ-GALLARDO, SIMON LLOYD and ED MANUEL: The Transmission of Macroprudential Policy in the Tails: Evidence from a Narrative Approach.

2520 ALICIA AGUILAR: Beyond fragmentation: unraveling the drivers of yield divergence in the euro area.

2521 RUBÉN DOMÍNGUEZ-DÍAZ and DONGHAI ZHANG: The macroeconomic effects of unemployment insurance extensions: A policy rule-based identification approach.

2522 IRMA ALONSO-ALVAREZ, MARINA DIAKONOVA and JAVIER J. PÉREZ: Rethinking GPR: The sources of geopolitical risk.

2523 ALBERTO MARTÍN, SERGIO MAYORDOMO and VICTORIA VANASCO: Banks vs. Firms: Who Benefits from Credit Guarantees?

2524 SUMIT AGARWAL, SERGIO MAYORDOMO, MARÍA RODRÍGUEZ-MORENO and EMANUELE TARANTINO: Household Heterogeneity and the Lending Channel of Monetary Policy.

2525 DIEGO BONELLI, BERARDINO PALAZZO, and RAM YAMARTHY: Good inflation, bad inflation: implications for risky asset prices.

2526 STÉPHANE BONHOMME and ANGELA DENIS: Fixed Effects and Beyond. Bias Reduction, Groups, Shrinkage and Factors in Panel Data.

2527 ÁLVARO FERNÁNDEZ-GALLARDO and IVÁN PAYÁ: Public debt burden and crisis severity.

2528 GALO NUÑO: Three Theories of Natural Rate Dynamics.

2529 GALO NUÑO, PHILIPP RENNER and SIMON SCHEIDEGGER: Monetary policy with persistent supply shocks.

2530 MIGUEL ACOSTA-HENAO, MARÍA ALEJANDRA AMADO, MONTSERRAT MARTÍ and DAVID PÉREZ-REYNA: Heterogeneous UIPDs across Firms: Spillovers from U.S. Monetary Policy Shocks.

2531 LUIS HERRERA and JESÚS VÁZQUEZ: Learning from news.

2532 MORTEZA GHOMI, JOCHEN MANKART, RIGAS OIKONOMOU and ROMANOS PRIFTIS: Debt maturity and government spending multipliers.

2533 MARINA DIAKONOVA, CORINNA GHIRELLI and JAVIER J. PÉREZ: Political polarization in Europe.

2534 NICOLÁS FORTEZA and SERGIO PUENTE: Measuring non-workers' labor market attachment with machine learning.

2535 GERGELY GANICS and LLUC PUIG CODINA: Simple Tests for the Correct Specification of Conditional Predictive Densities.