

Conformal Quantile Nowcasting^{*}

Martin Fankhauser[†]

October 11, 2026

Abstract

Quantile nowcasts of GDP growth are central to monitoring macroeconomic tail risk, but producing them often requires computationally demanding models or bootstrap procedures. We propose *conformal quantile nowcasting*, a computationally inexpensive, model-agnostic approach that adds empirical quantiles of past real-time forecast errors to current point nowcasts. Obtaining multiple quantiles requires no additional model estimation or simulation, making the computational cost essentially that of point nowcasting. The procedure is a one-sided version of split conformal prediction, with finite-sample calibration under exchangeability, approximate calibration under weak dependence, and non-crossing quantiles by construction. In simulations, it is robust to many regressors and outliers, both of which distort residual bootstraps. In weekly nowcasts of the 5% quantile of US GDP growth, conformalized MIDAS performs comparably to a Bayesian mixed-frequency model with stochastic volatility, while a MIDAS bootstrap is miscalibrated. For daily stock returns, the method performs on par with a dynamic multiple quantile model.

Keywords: Conformal Prediction, Nowcasting, Quantile Forecasts, Tail Risk, Forecast Errors, Bootstrap

JEL Codes: C14, C53, E17, E37

^{*}Thanks to Massimiliano Marcellino, Todd Clark and Andrea Carriero as well as Alessandra Luati for providing me with the data and comments. In addition, thanks to Pamela Giustinelli, Sonia Petrone, Knut Are Aastveit and participants at the 3rd International Econometrics PhD Conference 2023 at Erasmus University Rotterdam, the Methods in Time Series Econometrics summer school 2023 in Bertinoro and the Virtual Workshop for Junior Researchers in Time Series 2024 for helpful discussions. Finally, thanks to Kaspar Wüthrich and Rina Foygel Barber for helpful inputs via email. All errors are my own.

[†]Bocconi University, Milan, Italy. Email: martin.fankhauser@unibocconi.it

1 Introduction

Policy institutions increasingly monitor not only the most likely path of economic activity but also its downside risks. Quantile forecasts of GDP growth, such as the 5% quantile emphasized by the growth-at-risk literature (Adrian, Boyarchenko, and Giannone 2019; Adrian, Grinberg, et al. 2022; Brownlees and Souza 2021), have become a standard summary of these risks, and a recent literature produces them in real time from high-frequency data (Lima, Meng, and Godeiro 2020; Carriero, Clark, and Marcellino 2022). The leading approach to such *quantile nowcasts* combines a mixed-frequency model for the conditional mean with stochastic volatility. Carriero, Clark, and Marcellino (2022), henceforth CCM2022, show that a Bayesian mixed-frequency model with stochastic volatility (BMF-SV) delivers more accurate quantile nowcasts than the same model without stochastic volatility (BMF), mixed-data sampling (MIDAS) regressions and quantile regressions, although adding stochastic volatility does not improve point nowcasts (see also Clark 2011; Carriero, Clark, and Marcellino 2015; Carriero, Clark, and Marcellino 2016). Stochastic volatility, however, comes at a cost: many latent states must be estimated by simulation, which makes real-time estimation slow, and it is not obvious how to add it to models such as MIDAS regressions that are estimated by non-linear least squares. Residual bootstraps for MIDAS models (Aastveit, Foroni, and Ravazzolo 2017) are an alternative, but they require re-estimating the model for every bootstrap replication and, as we show below, can be badly miscalibrated when the model contains many regressors.

This paper proposes a simple alternative, which we call *conformal quantile nowcasting*. Given any point nowcasting model, the α -quantile nowcast is the current point nowcast plus the empirical α -quantile of the model’s past real-time nowcast errors (Algorithm 1). The procedure only requires the point nowcasts that are produced anyway, it can be wrapped around any model, from MIDAS regressions to machine-learning methods, and estimating several quantiles at once cannot produce quantile crossing. Constructing intervals from the empirical distribution of past forecast errors has a long tradition in applied forecasting

(Williams and Goodman 1971; Lee and Scholtes 2014; Knüppel 2014). We show that, when applied to *signed* errors and one quantile level at a time, this practice is a one-sided version of split conformal prediction (Vovk, Gammerman, and Shafer 2005; Lei, G’Sell, et al. 2018). This connection supplies what the heuristic lacks: explicit conditions under which the resulting quantile nowcasts are calibrated, and a precise statement of what they target.

The paper makes three contributions. First, we formalize conformal quantile nowcasting for a real-time environment with publication lags and multiple quantile levels, and we state the properties it inherits from conformal prediction (Section 2). If the nowcast errors are exchangeable, the probability that the outcome falls below the α -quantile nowcast is within $2/(n+1)$ of α , where n is the number of past errors, and quantile nowcasts at different levels never cross. For weakly dependent (β -mixing) data, the results of Oliveira et al. (2024) imply that these guarantees hold approximately. We also make explicit what the procedure targets: marginal quantiles, which coincide with conditional quantiles when the nowcast errors are independent of the predictors. Using signed rather than absolute errors matters, because standard two-sided conformal intervals control only the total miscoverage and can misallocate it between the tails when errors are skewed (Linusson, Johansson, and Löfström 2014).

Second, we compare conformal quantile forecasts with residual bootstraps for autoregressions (Pan and Politis 2016) in Monte Carlo simulations (Section 3). Without exogenous regressors, both approaches are close to nominal coverage, with the bootstrap more accurate in very small samples and for extreme quantiles; the small-sample behavior of the conformal quantiles is exactly what the finite-sample theory predicts. With many exogenous regressors, the non-studentized bootstrap overcovers, while conformal quantiles adapt to the dimension of the problem. The studentized bootstrap repairs the dimensionality problem but overcovers when the sample contains outliers, whereas conformal quantiles, which depend on the errors only through an order statistic, are unaffected by the outlier magnitude.

Third, we show in two real-time applications that this simple wrapper is competitive with state-of-the-art quantile models. In the main application (Section 4), we nowcast

the 5% quantile of US GDP growth at a weekly frequency in the setting of CCM2022. A conformalized MIDAS regression estimated by parameter profiling (Ghysels and Qian 2019) attains empirical coverage that is not significantly different from the nominal level at the 5% level in any of the 15 weeks of the nowcasting cycle, and quantile scores that are lower than those of the BMF-SV in 9 of 15 weeks, although none of these differences is statistically significant. The MIDAS bootstrap of Aastveit, Foroni, and Ravazzolo (2017), in contrast, produces significantly too few exceedances in every week. In a second application (Section 5), we forecast all percentiles of daily Microsoft stock returns and compare with the dynamic multiple quantile (DMQ) model of Catania and Luati (2023), which is designed to avoid quantile crossing. A conformalized autoregression yields quantile losses within 3% of the DMQ for every decile and comparable backtesting results, while it avoids quantile crossing by construction and is cheap enough to be re-estimated every day.

The paper relates to several strands of the literature. Conformal prediction converts point predictions into prediction sets with finite-sample coverage guarantees under exchangeability (Vovk, Gammerman, and Shafer 2005; Lei, G’Sell, et al. 2018); see Angelopoulos and Bates (2023) and Fontana, Zeni, and Vantini (2023) for reviews. One-sided conformal sets based on signed scores were introduced by Linusson, Johansson, and Löfström (2014) to control the coverage of each tail and appear in Romano, Patterson, and Candès (2019) and Wang and Politis (2021); Vovk, Shen, et al. (2019) show how conformal methods deliver entire predictive distributions. Our focus differs in that we use one-sided conformal sets as quantile forecasts built from point forecasts alone, whereas conformalized quantile regression (Romano, Patterson, and Candès 2019) and distributional conformal prediction (Chernozhukov, Wüthrich, and Zhu 2021b) require a quantile or distribution model as input. A growing literature extends conformal prediction to dependent data (Chernozhukov, Wüthrich, and Zhu 2018; Gibbs and Candès 2021; Gibbs and Candès 2024; Barber et al. 2023; Xu and Xie 2023; Oliveira et al. 2024), which we review in Section 2.4. Finally, we contribute to the literature on density and quantile nowcasting (Aastveit, Gerdrup, et al. 2014; Aastveit,

Foroni, and Ravazzolo 2017; Lima, Meng, and Godeiro 2020; Carriero, Clark, and Marcellino 2022) by providing a model-agnostic and computationally light way to obtain calibrated quantile nowcasts.

The remainder of the paper is organized as follows. Section 2 presents the method, its finite-sample properties and its relation to conformal methods for dependent data. Section 3 reports the Monte Carlo evidence, Section 4 the nowcasting application and Section 5 the multiple quantile application. Section 6 concludes. Proofs are collected in Appendix A and additional results in Appendix B.

2 Conformal quantile nowcasts

2.1 Setting and algorithm

Let $y_t \in \mathbb{R}$ denote the target variable in period t , for instance quarterly GDP growth, and let $\mathcal{I}_t$ denote the information set available when the nowcast of y_t is made. In nowcasting, $\mathcal{I}_t$ contains past values of y that have already been published and high-frequency indicators released up to the nowcast date. A point nowcasting method $\mathcal{A}$, for instance a MIDAS regression, a Bayesian mixed-frequency model or a machine-learning algorithm, is estimated on $\mathcal{I}_t$ and produces the point nowcast $\hat{y}_t$. We place no restriction on $\mathcal{A}$; in particular, it can be re-estimated in every period.

Once y_t is published, the *signed* real-time nowcast error

$$e_t = y_t - \hat{y}_t \tag{1}$$

becomes available. Because of publication lags, the error for period s is known at the time of the nowcast for period t only if $s \leq t - \ell$, where $\ell \geq 1$ is the publication lag measured in periods of the target variable. For a window length $w \in \mathbb{N} \cup \{\infty\}$, the set of past errors

available when nowcasting y_t is

$$\mathcal{E}_t = \{e_s : t - \ell - w < s \leq t - \ell, s \geq s_0\}, \quad n_t = |\mathcal{E}_t|, \quad (2)$$

where s_0 is the first period for which an error is recorded and $w = \infty$ corresponds to an expanding window.

For a finite set $\mathcal{E}$ of n numbers with order statistics $e_{(1)} \leq \dots \leq e_{(n)}$, let $\hat{Q}_\alpha(\mathcal{E})$ denote its empirical α -quantile. We consider two versions. The *conservative* version sets $\hat{Q}_\alpha(\mathcal{E}) = e_{(\lfloor \alpha(n+1) \rfloor)}$, with the convention $e_{(0)} = -\infty$. The *interpolated* version, which is the default in R (type 7 in Hyndman and Fan 1996) and is used in all computations of this paper, sets $\hat{Q}_\alpha(\mathcal{E}) = e_{(j)} + \gamma(e_{(j+1)} - e_{(j)})$ with $h = (n-1)\alpha + 1$, $j = \lfloor h \rfloor$ and $\gamma = h - j$.

Definition 1 (Conformal quantile nowcast). *The conformal α -quantile nowcast of y_t is*

$$\hat{q}_t(\alpha) = \hat{y}_t + \hat{Q}_\alpha(\mathcal{E}_t). \quad (3)$$

Algorithm 1 summarizes the procedure. In the empirical applications, $\ell = 1$ unless the previous period's outcome has not yet been published at the nowcast date (see Section 4), and the window is either expanding ($w = \infty$) or rolling.

Algorithm 1 Conformal quantile nowcasting

Input: quantile levels $0 < \alpha_1 < \dots < \alpha_K < 1$; point nowcasting method $\mathcal{A}$; publication lag $\ell \geq 1$; window length w ; first error period s_0 and first forecast period $t_0 > s_0$

```

1: for  $t = s_0, s_0 + 1, \dots, T$  do
2:   Estimate  $\mathcal{A}$  on the information set  $\mathcal{I}_t$  and compute the point nowcast  $\hat{y}_t$ 
3:   if  $t \geq t_0$  then
4:     Collect the published errors  $\mathcal{E}_t = \{e_s : t - \ell - w < s \leq t - \ell, s \geq s_0\}$ 
5:     for  $k = 1, \dots, K$  do
6:        $\hat{q}_t(\alpha_k) \leftarrow \hat{y}_t + \hat{Q}_{\alpha_k}(\mathcal{E}_t)$  ▷ empirical quantile of signed errors
7:     end for
8:   end if
9:   Once  $y_t$  is published, record the signed error  $e_t = y_t - \hat{y}_t$ 
10: end for
```

Output: quantile nowcasts $\hat{q}_t(\alpha_1) \leq \dots \leq \hat{q}_t(\alpha_K)$ for $t = t_0, \dots, T$

We refer to Algorithm 1 as CP1 and add the name of the point nowcasting method where needed, for instance CP1-MIDAS. As a benchmark we also consider the lower end of the standard two-sided split conformal interval, which uses absolute errors as conformity scores, $\hat{q}_t^{\text{CP2}}(\alpha) = \hat{y}_t - \hat{Q}_{1-2\alpha}(\{|e_s| : e_s \in \mathcal{E}_t\})$; we refer to this version as CP2. The computational cost of CP1 is that of the point nowcasts plus sorting at most n_t numbers per period. No additional model has to be estimated, and the point nowcasting model does not have to be re-estimated on resampled data as in the bootstrap.

2.2 Finite-sample properties

The conformal quantile nowcast is the lower end of a one-sided split conformal prediction set based on the signed score $y - \hat{y}$ (Linusson, Johansson, and Löfström 2014). Its properties therefore follow from standard conformal arguments (Vovk, Gammerman, and Shafer 2005; Lei, G'Sell, et al. 2018; Romano, Patterson, and Candès 2019), which we restate for the quantile nowcast in (3).

Proposition 1 (Finite-sample calibration). *Fix a forecast period t and suppose that the $n + 1$ errors in $\mathcal{E}_t \cup \{e_t\}$ are exchangeable and almost surely distinct, where $n = n_t$. Let $e_{(1)} < \dots < e_{(n)}$ denote the order statistics of $\mathcal{E}_t$.*

(i) *For every $k \in \{1, \dots, n\}$, $P(y_t \leq \hat{y}_t + e_{(k)}) = k/(n + 1)$.*

(ii) *For the conservative version of $\hat{Q}_\alpha$ and $\alpha \geq 1/(n + 1)$,*

$$\alpha - \frac{1}{n + 1} < P(y_t \leq \hat{q}_t(\alpha)) \leq \alpha.$$

(iii) *For the interpolated version of $\hat{Q}_\alpha$, with $j = \lfloor (n - 1)\alpha \rfloor + 1$,*

$$\frac{j}{n + 1} \leq P(y_t \leq \hat{q}_t(\alpha)) \leq \frac{j + 1}{n + 1}, \quad \text{so that} \quad \left| P(y_t \leq \hat{q}_t(\alpha)) - \alpha \right| \leq \frac{2}{n + 1}.$$

(iv) *For both versions and any $\alpha < \alpha'$, $\hat{q}_t(\alpha) \leq \hat{q}_t(\alpha')$ for every realization of the data.*

The proof is given in Appendix A. Parts (i)–(iii) say that the conformal quantile nowcast is calibrated in finite samples: the frequency with which the outcome falls below $\hat{q}_t(\alpha)$ is within a known distance of α , without any assumption on the distribution of the errors or on the point nowcasting method. Part (iv) says that quantile nowcasts at different levels never cross, because they share the same point nowcast and differ only through the monotone empirical quantile function. Four remarks clarify the scope of these results.

Remark 1 (When are the errors exchangeable?). Exchangeability holds, for instance, if the data are independent and identically distributed and the point nowcasting method is estimated once on a separate training sample, which is the classical split conformal setting (Vovk, Gammerman, and Shafer 2005; Lei, G’Sell, et al. 2018). Full conformal prediction preserves exchangeability for methods that are re-estimated on the augmented sample, provided the estimator treats the observations symmetrically, which rules out, for instance, estimators that weight observations by their position in time (Vovk, Gammerman, and Shafer 2005; Barber et al. 2023). In nowcasting, neither condition holds exactly: the data are serially dependent and the point nowcasting model is re-estimated every period on a growing information set. Section 2.4 discusses the approximate validity of the procedure in this case, and Sections 3–5 assess it numerically.

Remark 2 (What is being forecast?). Proposition 1 concerns the *marginal* probability $P(y_t \leq \hat{q}_t(\alpha))$, averaged over the predictors. Conditional calibration, $P(y_t \leq \hat{q}_t(\alpha) \mid x_t) = \alpha$ for all values of the predictors x_t , cannot be guaranteed in finite samples without further assumptions (Lei and Wasserman 2014). Suppose $y_t = \mu(x_t) + \varepsilon_t$ with ε_t independent of x_t , and that the point nowcast is consistent for $\mu(x_t)$. Then the conformal quantile nowcast converges to $\mu(x_t) + Q_\alpha(\varepsilon_t)$, the conditional α -quantile of y_t . If instead the dispersion of ε_t varies with x_t or over time, the conformal quantile nowcast remains marginally calibrated but its shape around the point nowcast does not adapt to the conditional dispersion. This is precisely the feature that stochastic volatility and quantile regression models are designed to capture. The empirical question addressed in Section 4 is how much is lost, in terms of coverage and

quantile scores, by relying on marginal calibration around a good point nowcast.

Remark 3 (Number of errors and multiple quantiles). By part (ii), the conservative version can only target levels $\alpha \geq 1/(n+1)$: with fewer than $1/\alpha - 1$ past errors, a calibrated α -quantile nowcast does not exist without further assumptions. The interpolated version always returns a finite value, but by part (iii) its exceedance probability is at least $1/(n+1)$. The same logic applies to multiple quantiles: to obtain distinct forecasts of all percentiles, at least 100 past errors are required.

Remark 4 (One-sided versus two-sided scores). Standard conformal intervals use absolute errors $|e_s|$ as conformity scores. A $(1 - 2\alpha)$ interval then guarantees that the *total* probability outside the interval is approximately 2α , but it does not control how this probability is split between the two tails. If the errors are skewed, the probability below the lower end can be anywhere between 0 and 2α . Signed errors guarantee calibration for each tail separately (Linusson, Johansson, and Löfström 2014). The cost is that a one-sided interval with the same nominal level is wider. This cost is immaterial when the object of interest is a quantile.

2.3 Relation to other conformal approaches to quantiles

The idea of constructing intervals from the empirical distribution of past forecast errors is old (Williams and Goodman 1971) and is still widely used, for instance to construct fan charts (Lee and Scholtes 2014; Knüppel 2014). Algorithm 1 differs from these approaches in two respects. It treats each tail separately through signed errors, and it computes the errors in real time, so that each error reflects the information and the estimated model available when the corresponding nowcast was made. Viewing the procedure as conformal prediction provides the finite-sample guarantees of Proposition 1 and the results for dependent data discussed below.

Within the conformal literature, two approaches produce quantile forecasts by construction. Conformalized quantile regression (Romano, Patterson, and Candès 2019) adjusts the output of a quantile regression, and distributional conformal prediction (DCP; Chernozhukov,

Wüthrich, and Zhu 2021b) conformalizes an estimated conditional distribution, for instance the posterior predictive distribution of a Bayesian model. Both approaches use quantile or distribution estimates as inputs and can therefore adapt to conditional heteroskedasticity, but they require such estimates to be available. Our approach targets the opposite situation, in which accurate point nowcasts exist but quantile or density models are costly to estimate. In the nowcasting application we include a one-sided version of DCP applied to the posterior predictive distribution of the BMF as a benchmark (DCP-BMF), which is valid for strongly mixing processes and also rules out quantile crossing (Chernozhukov, Wüthrich, and Zhu 2021b). More generally, Vovk, Shen, et al. (2019) show that the collection of one-sided conformal sets at all levels forms a predictive distribution; Algorithm 1 with many quantile levels is a split-conformal instance of this idea. Conformalizing Bayesian predictive distributions can also be viewed as a protective layer against misspecification (Fong and Holmes 2021), a direction we do not pursue further.

2.4 Beyond exchangeability

Time series data are not exchangeable, and a large literature extends conformal prediction to this case. These extensions typically require additional assumptions on the data generating process, deliver weaker guarantees, or both. We summarize the main approaches and their implications for quantile nowcasting.

For weakly dependent data, Oliveira et al. (2024) show that split conformal prediction remains approximately valid for stationary β -mixing processes, a class that includes, for instance, stationary ARMA and GARCH processes under standard conditions. The coverage penalty depends on the mixing coefficients and vanishes as the number of calibration scores grows. Their results hold for any conformity score, so they apply to the signed errors used in Algorithm 1. They are stated for a model estimated once on a training sample, and re-estimating the model as new data arrive is discussed only informally. Because these results require no change to the algorithm, they justify applying Algorithm 1 directly to several

quantile levels, which preserves the monotonicity in Proposition 1(iv). We follow this route and assess its performance numerically.

A second strand builds on Gibbs and Candès (2021). Adaptive conformal inference (ACI) treats the level of the empirical quantile as a parameter that is updated online: the level is moved outward after a miss and inward otherwise. Instead of modeling distribution shifts in the high-dimensional joint distribution of the data (as, e.g., Cauchois et al. 2023), the problem reduces to tracking a scalar. The key tuning parameter is the step size of the update. Gibbs and Candès (2024) aggregate over a grid of step sizes with exponential reweighting (fully adaptive conformal inference, FACI) and obtain coverage guarantees over arbitrary time intervals. Zaffran et al. (2022) and Bastani et al. (2022) provide related proposals with long-run guarantees, Feldman et al. (2023) extend the idea to risk control, and Angelopoulos, Candès, and Tibshirani (2023) and Xu and Xie (2023) additionally forecast the quantile to be used in the next period. These methods require no assumptions on the data generating process, but the levels are updated separately for each quantile, so monotonicity across quantile levels is not guaranteed. We include FACI wrapped around CP1 as a benchmark.

A third strand reweights the conformity scores. Tibshirani et al. (2019) use a likelihood ratio to accommodate covariate shift, and Barber et al. (2023) use fixed, data-independent weights that accommodate gradual distribution shifts. Their guarantees deteriorate with the size of the violation of exchangeability. Barber et al. (2023) also handle non-symmetric estimators by randomly swapping the test point with a calibration point. In autoregressive models this swapping can harm predictive accuracy, because recent observations are the most informative. The sequential approach of Xu and Xie (2023) can be viewed as a data-driven choice of such weights. A different line of work builds on block permutations of conformity scores (Chernozhukov, Wüthrich, and Zhu 2018; Chernozhukov, Wüthrich, and Zhu 2021a). Xu and Xie (2021) use this idea for a bootstrap aggregation scheme related to the jackknife+ (Barber et al. 2021), and Jensen, Bianchi, and Anfinson (2022) extend conformalized quantile regression to time series. Finally, Stankeviciute, Alaa, and Schaar (2021) and Lin, Trivedi,

and Sun (2022) consider multi-horizon and cross-sectional time series, respectively.

Based on the results of Oliveira et al. (2024) and on its performance relative to these alternatives in our applications, we use Algorithm 1 as the main method whenever only point nowcasts are available. As benchmarks, we report FACI wrapped around CP1, because it requires no assumptions on the data generating process, and, in the nowcasting application, DCP applied to the BMF. All approaches above can in principle be used to produce quantile forecasts, but only some of them guarantee non-crossing quantiles.

3 Monte Carlo simulations

This section compares conformal quantile forecasts with bootstrap quantile forecasts for linear autoregressions, with and without exogenous regressors, and compares conformal and DMQ forecasts of multiple quantiles. The designs mirror the two empirical applications.

For the purely autoregressive case, we compare with the forward bootstrap of Pan and Politis (2016). For the models with exogenous regressors, we apply the same procedure and treat the exogenous regressors as fixed, as is common in the forecasting literature (e.g., Aastveit, Foroni, and Ravazzolo 2017). We consider the non-studentized forward bootstrap with predictive residuals (FpB) and its studentized counterpart with fitted residuals (FsB), Algorithms 3.2 and 3.3 in Pan and Politis (2016).¹ CP1 denotes Algorithm 1 with an expanding window of one-step-ahead out-of-sample errors from the re-estimated autoregression, and CP2 the lower end of the corresponding two-sided interval. In each replication, a sample of size T is generated and used to construct the forecast for period $T + 1$. To measure coverage conditional on the sample, we draw 1,000 realizations of y_{T+1} for each simulated sample and record the fraction that falls below the forecast quantile. We report averages over 1,000 replications.²

¹We do not consider the studentized version with predictive residuals because Pan and Politis (2016) report a similar performance to the two versions considered.

²The bootstrap uses 1,000 bootstrap replications. Conformal prediction guarantees marginal rather than conditional coverage without further assumptions on the data generating process (Lei and Wasserman 2014);

In the first design, we follow Pan and Politis (2016) and simulate from an AR(1) model with coefficient $\phi = 0.5$, with errors drawn either from a standard normal distribution or from a χ^2 distribution with 5 degrees of freedom, shifted to have mean zero so that the model for the conditional mean is correctly specified.³ Errors for CP1 are collected from period 20 onward, so CP1 is based on $n = T - 20$ errors.

The top panel of Table 1 reports the exceedance frequencies for the 1%, 5% and 10% quantiles with $T = 50, 100$ and 200 and normal errors; the last three rows report the coverage of two-sided prediction intervals with nominal level $1 - 2\alpha$. The bootstraps are closer to the nominal level than the conformal methods for small samples and extreme quantiles, in line with the simulation results of Wang and Politis (2021). This small-sample behavior of CP1 is the one predicted by Proposition 1. For $T = 50$, CP1 uses $n = 30$ errors, and part (iii) of the proposition implies an exceedance probability between $1/31 \approx 0.032$ and $2/31 \approx 0.065$ for $\alpha = 0.01$; the simulated value is 0.036 . In fact, all nine CP1 entries in the top panel lie within the bounds of part (iii), even though the re-estimated errors are only approximately exchangeable. The deviation from the nominal level for $\alpha = 0.01$ in small samples is therefore a consequence of the small number of past errors (Remark 3) and vanishes as T grows.

The bottom panel repeats the exercise with skewed χ^2 errors, so that the point forecast makes asymmetric mistakes. CP1 performs as with normal errors, whereas CP2 is heavily miscalibrated: its exceedance frequency is close to zero for $\alpha = 0.01$ and about 0.01 – 0.02 for $\alpha = 0.05$. This is the misallocation of tail probabilities described in Remark 4. The bootstraps again perform best, relative to the conformal methods, in small samples and for extreme quantiles.

The second design, reported in Table 2, augments the autoregression with $K = 10, 25$ and 50 exogenous regressors drawn from a standard normal distribution, with $T = 100$, so that

see also Angelopoulos and Bates (2023). Under additional assumptions some notion of conditional coverage can be achieved (Oliveira et al. 2024; Chernozhukov, Wüthrich, and Zhu 2021b), which is why Wang and Politis (2021) focus their comparison of the model-free bootstrap on DCP.

³Oliveira et al. (2024) show in simulations that the split conformal algorithm works well even for highly persistent processes. In our simulations, the autoregressive parameter was not crucial as long as the process remains stationary.

the ratio of regressors to observations is $\kappa = K/T = 0.10, 0.25$ and 0.50 .⁴ This design mimics a MIDAS regression with many high-frequency regressors and a lagged dependent variable, as bootstrapped by Aastveit, Foroni, and Ravazzolo (2017). The non-studentized bootstrap overcovers, and increasingly so as κ grows; at $\kappa = 0.5$ its exceedance frequency for the 10% quantile is 0.047. A similar failure of the bootstrap in high dimensions is documented by El Karoui and Purdom (2018) for inference on regression coefficients (see Chernozhukov, Chetverikov, et al. 2023, for a review). Studentizing the bootstrap with fitted residuals largely solves the problem. The conformal methods are much less affected by the dimension than the non-studentized bootstrap, although they become somewhat conservative as κ increases. A plausible explanation is that the errors collected early in the sample come from models estimated on few observations relative to the number of regressors. These errors are larger than those of the final model, which inflates the empirical quantile. Overall, Table 2 shows that in high-dimensional problems either a studentized bootstrap or conformal quantiles are approximately valid.

The studentized bootstrap, however, is sensitive to outliers. Table 3 repeats the AR(1) design with normal errors and $T = 100$, but replaces one error in the sample by a draw from $-|N(\mu_o, 0.5)|$, so that 1% of the sample is an outlier. Future errors are drawn from the corresponding mixture distribution, so that the sample and the future observations come from the same distribution. CP1 remains close to the nominal level for all values of μ_o , whereas both bootstraps are miscalibrated, and the studentized bootstrap increasingly so as the outlier magnitude grows. The reason is that the studentized bootstrap scales the residuals by their sample standard deviation, which is driven by the outlier, whereas the conformal quantile depends on the errors only through an order statistic: a single outlying error can shift $\hat{Q}_\alpha(\mathcal{E}_t)$ by at most one position in the ordered sample, whatever its magnitude. This finding, together with the failure of the non-studentized bootstrap in high dimensions, is relevant for the nowcasting application, where the GDP growth series contains several

⁴To ensure that the least squares estimator is unique, we start collecting errors for CP1 in periods 20, 45 and 60, respectively.

large negative observations associated with recessions and the MIDAS model contains many regressors. As shown in Section 4, none of the bootstrap procedures attains coverage close to the nominal level there.

Finally, to study the joint forecasting of multiple quantiles, we simulate $T = 50, 100$ and 200 observations from the DMQ model of Catania and Luati (2023). We use the same sample to estimate the DMQ and to construct CP1 forecasts of the nine deciles of the one-step-ahead distribution, based on an AR(4) point forecast and errors collected from period 20 onward.⁵ Table 4 reports quantile losses of CP1 relative to the DMQ, averaged over 500 replications, so that values below one favor CP1. Even though the data are generated by the DMQ, CP1 performs similarly to the DMQ in finite samples. Its loss is at most 6% higher, and the difference tends to be smaller for higher quantiles, a pattern that also appears in the empirical application of Section 5.

4 Nowcasting US GDP tail risk

Nowcasting is the prediction of the present, the very near future and the very recent past. The main challenge is to exploit indicators that are available at a daily, weekly or monthly frequency to predict a quarterly variable such as GDP growth before it is published (see Bańbura et al. 2013, for a review). Following CCM2022, we produce nowcasts in each of 15 weeks per quarter: four weeks in each month of the quarter and the first three weeks of the following quarter, before the first estimate of GDP is released in the fourth week.

4.1 Models

Nowcasting models can be broadly divided into full-system and partial approaches. Full-system approaches specify a joint model for all low- and high-frequency variables, for instance a mixed-frequency VAR (Schorfheide and Song 2015). Partial approaches estimate a single

⁵We forecast deciles rather than percentiles to reduce the computational cost of the DMQ.

equation for the low-frequency target in which the high-frequency variables are aligned with the low frequency. With monthly indicators and a quarterly target, for example, each indicator enters with one regressor per month of the quarter, so that even a moderate number of indicators produces many regressors. This parameter proliferation is addressed either by Bayesian shrinkage, as in the Bayesian mixed-frequency models with and without stochastic volatility of Carriero, Clark, and Marcellino (2015) (BMF-SV and BMF), or by restricting the lag coefficients to a parametric polynomial, as in mixed-data sampling (MIDAS) regressions (Ghysels, Santa-Clara, and Valkanov 2004).

CCM2022 show that the BMF and the BMF-SV produce point nowcasts of similar accuracy, but that the BMF-SV produces better quantile and density nowcasts at a much higher computational cost. Conformal quantile nowcasting suggests a way around this trade-off: combine a fast and accurate point nowcast with Algorithm 1. We apply the algorithm to a MIDAS regression with many high-frequency regressors whose polynomial lag parameters are restricted to be common across indicators, estimated by parameter profiling (Ghysels and Qian 2019) (CP1-MIDAS), and compare it with the BMF-SV. The specifications of the BMF and the BMF-SV, including the priors, follow CCM2022.

We also report five alternatives: CP2-MIDAS, the lower end of the two-sided conformal interval; FACI-MIDAS, the FACI algorithm of Gibbs and Candès (2024) wrapped around CP1-MIDAS, which allows for arbitrary distribution shifts; DCP-BMF, the one-sided DCP of Chernozhukov, Wüthrich, and Zhu (2021b) applied to the posterior predictive distribution of the BMF; BOOT-MIDAS, the MIDAS bootstrap of Aastveit, Foroni, and Ravazzolo (2017);⁶ and the two forward bootstraps of Section 3 (Appendix B.2).

4.2 Data and real-time design

We use the data of CCM2022, organized as a weekly calendar of data releases, and the set of indicators that CCM2022 refer to as the *base macro-finance* set. It contains

⁶Aastveit, Foroni, and Ravazzolo (2017) consider models with a single high-frequency regressor.

six monthly non-financial indicators: payroll employment, industrial production, real retail sales (nominal sales deflated by the CPI), housing starts, the manufacturing index of the Institute for Supply Management (ISM) and initial claims for unemployment insurance. It also contains four financial indicators: the Chicago Fed’s National Financial Conditions Index (NFCI), the S&P 500 index, the term spread between the 10-year and 1-year Treasury yields (constant maturity), and the credit spread between Moody’s Baa corporate yield and the 10-year Treasury yield. Some of these variables are also available weekly. Depending on the week, they enter at the weekly or the monthly frequency, and weekly observations are sometimes aggregated; see Table 2 of CCM2022 for details. Our aim is methodological, and the specifications only need to ensure comparability with CCM2022.

All results are real-time and out-of-sample. Nowcast errors are recorded from 1985Q1 onward, and the evaluation sample runs from 1989Q4 to 2019Q3 (120 quarters). The outcome is the second release of GDP growth, taken from the CCM2022 replication files. Algorithm 1 is applied separately for each week of the nowcasting cycle, using the errors of the nowcasts made in the same week of previous quarters and an expanding window ($w = \infty$). The publication lag follows the data calendar. In weeks 1–3, the previous quarter’s GDP has not yet been released, so the most recent available error is for two quarters earlier ($\ell = 2$). From week 4 onward, the previous quarter’s error is available ($\ell = 1$). The COVID-19 period is excluded from the evaluation and discussed separately.

4.3 Evaluation

Let $\hat{q}_t$ denote the 5% quantile nowcast for quarter t in a given week. We report the empirical exceedance frequency $\frac{1}{P} \sum_t \mathbf{1}\{y_t \leq \hat{q}_t\}$ over the P evaluation quarters and test whether it equals the nominal level of 5% with a two-sided t -test. With 120 evaluation quarters, about six exceedances are expected, so these tests have limited power. We therefore

also report the quantile score

$$\text{QS}_\alpha(\hat{q}_t, y_t) = (\mathbf{1}\{y_t \leq \hat{q}_t\} - \alpha) (\hat{q}_t - y_t), \quad (4)$$

which is a consistent scoring function for the α -quantile and therefore rewards both calibration and sharpness (Gneiting 2011). We compare average quantile scores with those of the BMF-SV using the test of Diebold and Mariano (1995).

4.4 Results

Table 5 reports exceedance frequencies of the 5% quantile nowcasts for odd weeks; Table 8 in Appendix B.1 reports even weeks. The exceedance frequencies of the BMF-SV and of CP1-MIDAS are not statistically different from the nominal level in any odd week. In the even weeks, the only rejection for CP1-MIDAS is at the 10% level in week 2. CP2-MIDAS, in contrast, has significantly too few exceedances in several weeks, which indicates that the MIDAS nowcast errors are asymmetric (Remark 4). Wrapping FACI around CP1-MIDAS worsens the results, which suggests that non-exchangeability is not a first-order issue in this application. The bootstrap and DCP-BMF perform worst. BOOT-MIDAS has far too few exceedances in every week, as predicted by the simulation evidence, because of the dimension of the MIDAS model, the outliers in the sample, or both.⁷ DCP-BMF errs in the opposite direction, with exceedance frequencies around 10% in most weeks.

Table 6 reports quantile scores. The first row contains the average score of the BMF-SV, and all other rows report ratios relative to it, so that values below one indicate a better performance than the BMF-SV; Table 9 in Appendix B.1 reports even weeks. CP1-MIDAS has a lower quantile score than the BMF-SV in 9 of the 15 weeks, but none of the differences between the two is statistically significant. The bootstrap and DCP-BMF again perform

⁷We include many high-frequency regressors in one MIDAS model and restrict the polynomial parameters to be common across variables. In line with the outlier simulations, the two forward bootstraps of Section 3 do not solve the problem (Appendix B.2). Unreported simulations confirm that, without outliers, the studentized bootstrap works well for a range of data generating processes, despite the common polynomial parameters.

worst, and the bootstrap is significantly worse than the BMF-SV in most weeks. CP2-MIDAS and FACI-MIDAS are comparable to the BMF-SV on average but significantly worse in some weeks (and, for CP2-MIDAS, significantly better in week 2).

Figure 1 plots the real-time 5% quantile nowcasts of CP1-MIDAS, FACI-MIDAS and the BMF-SV for weeks 1, 5, 10 and 15. The conformal quantile nowcasts become smoother as the quarter progresses. In weeks 1 and 5 they differ considerably from the BMF-SV, whereas in weeks 10 and 15 the BMF-SV and the conformalized MIDAS model nearly coincide. This pattern is consistent with Remark 2: as more information arrives within the quarter, the point nowcast improves, and the time variation of the quantile nowcast is increasingly driven by the point nowcast rather than by changes in the dispersion of the errors. A notable exception is the financial crisis, for which the BMF-SV predicts much more extreme values in week 15 than the other models. The difference becomes more pronounced when the sample is extended to the COVID-19 period (Figure 2). The BMF-SV reacts strongly to the extreme observations and produces very low quantile nowcasts in weeks 1, 10 and 15, whereas the conformal quantile nowcasts do not. An outlier-robust specification would mitigate this behavior of the BMF-SV. The conformal approach is robust without such a modification, because extreme past errors affect the empirical quantile only through their rank.

Overall, conformal quantile nowcasts outperform the bootstrap of Aastveit, Foroni, and Ravazzolo (2017) in this high-dimensional application in terms of both coverage and quantile scores. They also perform similarly to the BMF-SV, which CCM2022 show to outperform quantile MIDAS regressions as well as frequentist and Bayesian quantile regressions. The main practical advantage is computational. The cost of Algorithm 1 is essentially that of the point nowcasting model, whereas the bootstrap re-estimates the point nowcasting model for every bootstrap replication and the stochastic volatility model adds a large number of latent states.

5 Forecasting multiple quantiles

When several quantiles are forecast separately, the forecasts may cross, which is incoherent (Bassett and Koenker 1982). The DMQ model of Catania and Luati (2023) is a recent score-driven model designed to forecast multiple quantiles while avoiding crossing. Its authors show that it outperforms the multiple-quantile CAViaR model (Engle and Manganelli 2004; White, Kim, and Manganelli 2015) and ARMA-GARCH models (Bollerslev 1986). This section compares conformal quantile forecasts, which cannot cross by Proposition 1(iv), with the DMQ in the empirical application of Catania and Luati (2023).

The target is the daily logarithmic return, in percentage points, of Microsoft Corporation stock from December 8, 2010, to November 15, 2018. We forecast all percentiles one day ahead and report only out-of-sample results. As in Catania and Luati (2023), the first 1,000 observations are used only for training. The DMQ is estimated once on these observations, and its quantiles are then updated with the most recent observations while its parameters are kept fixed. For the conformal methods, the point forecast is an AR(4) model re-estimated every day on an expanding window. The lag order is selected on the first 100 observations. Algorithm 1 uses a rolling window of the $w = 100$ most recent one-step-ahead errors with $\ell = 1$. We also wrap FACI around CP1. Re-estimating the AR(4) every day takes seconds on a laptop, whereas re-estimating the DMQ every day would be prohibitively expensive. The comparison therefore reflects what each method can feasibly do in real time. Because daily returns are nearly unpredictable in mean, CP1 is close in spirit to historical simulation with a rolling window. The comparison therefore also shows how much a carefully specified quantile model gains relative to a simple, model-free benchmark with finite-sample guarantees.

The top panel of Figure 3 displays the one-step-ahead forecasts of the 5%, 10%, ..., 95% quantiles from CP1, FACI and the DMQ. The DMQ produces smoother trajectories, reflecting its semi-parametric structure. The bottom panel reports p -values of the unconditional and conditional backtests of Kupiec (1995) and Christoffersen (1998). Under the null hypothesis, the indicator that the return falls below the forecast α_k -quantile is an i.i.d. Bernoulli variable

with success probability α_k . Following Catania and Luati (2023), we conduct the tests separately for each percentile. The unconditional test checks correct coverage on average, and the conditional test additionally checks the independence of exceedances over time. In the unconditional test, all methods perform well, with little evidence against the null, which is what Proposition 1 and the results of Oliveira et al. (2024) lead us to expect for CP1. Somewhat surprisingly, the null of correct conditional coverage is rejected only slightly more often for CP1 than for the DMQ. FACI adapts to local distribution shifts and achieves the best conditional coverage, but, unlike CP1 and the DMQ, it cannot guarantee monotonicity across quantile levels.

Table 7 reports aggregate quantile losses, defined as the sum of the percentile losses within each decile, relative to the DMQ, so that values below one favor the conformal methods. CP1 has a slightly lower loss than the DMQ for all but the first decile, with ratios between 0.973 and 1.003, and FACI has a lower loss in every decile. In terms of statistical significance, the 75% model confidence set of Hansen, Lunde, and Nason (2011) contains all three methods for the first five deciles, whereas for the remaining deciles it contains only FACI. Part of the gain relative to the DMQ comes from re-estimating the point forecasting model every day. However, the simulations in Section 3 show that, even when the data are generated by the DMQ and the DMQ is re-estimated in every replication, conformal quantile forecasts are only slightly worse than the DMQ in finite samples. Given the large sample, the similarity between the DMQ and the conformal forecasts is consistent with the DMQ being well specified for these data. Appendix B.3 shows how the results depend on the window length w .

6 Conclusion

This paper proposes conformal quantile nowcasting, which turns the point nowcasts of any model into quantile nowcasts by adding empirical quantiles of the model’s past real-time signed errors. The procedure is a one-sided version of split conformal prediction. It is

calibrated in finite samples when the errors are exchangeable and approximately calibrated for weakly dependent data. Quantile nowcasts at different levels cannot cross, and the procedure costs little more than the point nowcast itself.

In simulations, conformal quantile forecasts compare well with residual bootstraps for autoregressions. With many exogenous regressors, the non-studentized bootstrap fails, which highlights the need to adapt bootstrap procedures to high-dimensional problems; the MIDAS bootstrap of Aastveit, Foroni, and Ravazzolo (2017) was not designed for this case and is miscalibrated in our application. The studentized bootstrap performs well in high dimensions but is sensitive to outliers, whereas conformal quantiles are not. In real-time nowcasts of the 5% quantile of US GDP growth, a conformalized MIDAS regression performs similarly to the BMF-SV, a leading model for this task, and outperforms the bootstrap. When forecasting all percentiles of daily stock returns, a conformalized autoregression performs on par with the DMQ while avoiding quantile crossing by construction.

The approach has limitations. First, it is model-free and calibrates quantiles marginally. It complements rather than replaces models that describe the conditional distribution, such as the Bayesian mixed-frequency models of Carriero, Clark, and Marcellino (2015) and the quantile models of Engle and Manganelli (2004) and Catania and Luati (2023), which are needed to learn about the drivers of tail risk. Second, in very small samples conformal quantiles are less accurate than the bootstrap or correctly specified models, and the number of past errors limits the quantile levels that can be targeted: at least 100 errors are needed to obtain distinct forecasts of all percentiles. Third, the finite-sample guarantees rely on exchangeability, and the results for dependent data are approximate.

Despite these limitations, the results suggest that, at least in the applications considered here, non-exchangeability is not a first-order issue: unadjusted conformal quantile forecasts are well calibrated and competitive with considerably more complex models. We believe that the algorithmic view of statistical inference advocated by Vovk, Gammerman, and Shafer (2005) has many more applications in time series econometrics and forecasting.

References

- Aastveit, K., C. Foroni, and F. Ravazzolo (2017). “Density forecasts with MIDAS models”. In: *Journal of Applied Econometrics* 32.4, pp. 783–801.
- Aastveit, K., K. Gerdrup, et al. (2014). “Nowcasting GDP in real time: A density combination approach”. In: *Journal of Business and Economic Statistics* 32.1, pp. 48–68.
- Adrian, T., N. Boyarchenko, and D. Giannone (2019). “Vulnerable growth”. In: *American Economic Review* 109.4, pp. 1263–1289.
- Adrian, T., F. Grinberg, et al. (2022). “The term structure of growth-at-risk”. In: *American Economic Journal: Macroeconomics* 14.3, pp. 283–323.
- Angelopoulos, A. and S. Bates (2023). “Conformal prediction: A gentle introduction”. In: *Foundations and Trends in Machine Learning* 16.4, pp. 494–591.
- Angelopoulos, A., E. Candes, and R. Tibshirani (2023). “Conformal PID Control for Time Series Prediction”. In: *NeurIPS*.
- Bańbura, M. et al. (2013). “Now-casting and the real-time data flow”. In: *Handbook of economic forecasting* 2, pp. 195–237.
- Barber, R. F. et al. (2021). “Predictive inference with the jackknife+”. In: *Annals of Statistics* 49.1, pp. 486–507.
- (2023). “Conformal prediction beyond exchangeability”. In: *Annals of Statistics* 51.2, pp. 816–845.
- Bassett, G. and R. Koenker (1982). “An Empirical Quantile Function for Linear Models With iid Errors”. In: *Journal of the American Statistical Association* 77, pp. 407–415.
- Bastani, O. et al. (2022). “Practical adversarial multivalid conformal prediction”. In: *Advances in Neural Information Processing Systems* 35.
- Bollerslev, T. (1986). “Generalized autoregressive conditional heteroskedasticity”. In: *Journal of Econometrics* 31.3, pp. 307–327.
- Brownlees, C. and A. B. Souza (2021). “Backtesting global growth-at-risk”. In: *Journal of Monetary Economics* 118, pp. 312–330.

- Carriero, A., T. Clark, and M. Marcellino (2015). “Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility”. In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 178.4, pp. 837–862.
- (2016). “Common drifting volatility in large Bayesian VARs”. In: *Journal of Business and Economic Statistics* 34.3, pp. 375–390.
- (2022). “Nowcasting tail risk to economic activity at a weekly frequency”. In: *Journal of Applied Econometrics* 37.5, pp. 843–866.
- Catania, L. and A. Luati (2023). “Semiparametric modeling of multiple quantiles”. In: *Journal of Econometrics* 237.2.
- Cauchois, M. et al. (2023). “Robust validation: Confident predictions even when distributions shift”. In: *Journal of the American Statistical Association*.
- Chernozhukov, V., D. Chetverikov, et al. (2023). “High-dimensional data bootstrap”. In: *Annual Review of Statistics and Its Application* 10, pp. 427–449.
- Chernozhukov, V., K. Wüthrich, and Y. Zhu (2018). “Exact and robust conformal inference methods for predictive machine learning with dependent data”. In: *Proceedings of the 31st Conference on Learning Theory*. Vol. 75. Proceedings of Machine Learning Research, pp. 732–749.
- (2021a). “An exact and robust conformal inference method for counterfactual and synthetic controls”. In: *Journal of the American Statistical Association* 116.536, pp. 1849–1864.
- (2021b). “Distributional conformal prediction”. In: *Proceedings of the National Academy of Sciences* 118.48, e2107794118.
- Christoffersen, P. (1998). “Evaluating interval forecasts”. In: *International Economic Review* 39.4.
- Clark, T. (2011). “Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility”. In: *Journal of Business and Economic Statistics* 29.3, pp. 327–341.
- Diebold, F. X. and R. S. Mariano (1995). “Comparing predictive accuracy”. In: *Journal of Business and Economic Statistics* 13.3, pp. 253–263.

- El Karoui, N. and E. Purdom (2018). “Can We Trust the Bootstrap in High-dimensions? The Case of Linear Models”. In: *Journal of Machine Learning Research* 19, pp. 1–66.
- Engle, R. and S. Manganelli (2004). “CAViaR: Conditional autoregressive value at risk by regression quantiles”. In: *Journal of Business and Economic Statistics* 22.4.
- Feldman, S. et al. (2023). “Achieving Risk Control in Online Learning Settings”. In: *Transactions on Machine Learning Research* 11.
- Fong, E. and C. Holmes (2021). “Conformal bayesian computation”. In: *Advances in Neural Information Processing Systems* 34, pp. 18268–18279.
- Fontana, M., G. Zeni, and S. Vantini (2023). “Conformal prediction: a unified review of theory and new challenges”. In: *Bernoulli* 29.1, pp. 1–23.
- Ghysels, E. and H. Qian (2019). “Estimating MIDAS regressions via OLS with polynomial parameter profiling”. In: *Econometrics and statistics* 9, pp. 1–16.
- Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). *The MIDAS touch: Mixed data sampling regression models*. Working Paper 2004s-20. CIRANO.
- Gibbs, I. and E. Candes (2021). “Adaptive conformal inference under distribution shift”. In: *Advances in Neural Information Processing Systems* 34, pp. 1660–1672.
- Gibbs, I. and E. J. Candès (2024). “Conformal inference for online prediction with arbitrary distribution shifts”. In: *Journal of Machine Learning Research* 25.
- Gneiting, T. (2011). “Quantiles as optimal point forecasts”. In: *International Journal of Forecasting* 27.2, pp. 197–207.
- Hansen, P., A. Lunde, and J. Nason (2011). “The Model Confidence Set”. In: *Econometrica* 79.2, pp. 453–497.
- Hyndman, R. J. and Y. Fan (1996). “Sample quantiles in statistical packages”. In: *The American Statistician* 50.4, pp. 361–365.
- Jensen, V., F. Bianchi, and S. Anfinson (2022). “Ensemble Conformalized Quantile Regression for Probabilistic Time Series Forecasting”. In: *IEEE Transactions on Neural Networks and Learning Systems*.

- Knüppel, M. (2014). “Efficient estimation of forecast uncertainty based on recent forecast errors”. In: *International Journal of Forecasting* 30.2, pp. 257–267.
- Kuchibhotla, A. (2021). “Exchangeability, conformal prediction, and rank tests”. In: *arXiv preprint*.
- Kupiec, P. (1995). “Techniques for verifying the accuracy of risk measurement models”. In: *Journal of Derivatives* 3.2.
- Lee, Y. S. and S. Scholtes (2014). “Empirical prediction intervals revisited”. In: *International Journal of Forecasting* 30.2, pp. 217–234.
- Lei, J., M. G’Sell, et al. (2018). “Distribution-free predictive inference for regression”. In: *Journal of the American Statistical Association* 113.523, pp. 1094–1111.
- Lei, J. and L. Wasserman (2014). “Distribution-free prediction bands for non-parametric regression”. In: *Journal of the Royal Statistical Society: Series B* 76.1, pp. 71–96.
- Lima, L., F. Meng, and L. Godeiro (2020). “Quantile forecasting with mixed-frequency data”. In: *International Journal of Forecasting* 36.3, pp. 1149–1162.
- Lin, Z., S. Trivedi, and J. Sun (2022). “Conformal prediction with temporal quantile adjustments”. In: *Advances in Neural Information Processing Systems* 35, pp. 31017–31030.
- Linusson, H., U. Johansson, and T. Löfström (2014). “Signed-error conformal regression”. In: *Advances in Knowledge Discovery and Data Mining: 18th Pacific-Asia Conference*.
- Oliveira, R. I. et al. (2024). “Split conformal prediction and non-exchangeable data”. In: *Journal of Machine Learning Research* 25.225, pp. 1–38.
- Pan, L. and D. N. Politis (2016). “Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions”. In: *Journal of Statistical Planning and Inference* 177, pp. 1–27.
- Romano, Y., E. Patterson, and E. J. Candès (2019). “Conformalized quantile regression”. In: *Advances in Neural Information Processing Systems*. Vol. 32.
- Schorfheide, F. and D. Song (2015). “Real-time forecasting with a mixed-frequency VAR”. In: *Journal of Business and Economic Statistics* 33.3, pp. 366–380.

- Stankeviciute, K., A. Alaa, and M. van der Schaar (2021). “Conformal time-series forecasting”. In: *Advances in Neural Information Processing Systems* 34, pp. 6216–6228.
- Tibshirani, R. et al. (2019). “Conformal prediction under covariate shift”. In: *Advances in neural information processing systems* 32.
- Vovk, V., A. Gammerman, and G. Shafer (2005). *Algorithmic Learning in a Random World*. Springer.
- Vovk, V., J. Shen, et al. (2019). “Nonparametric predictive distributions based on conformal prediction”. In: *Machine Learning* 108, pp. 445–474.
- Wang, Y. and D. N. Politis (2021). *Model-free bootstrap and conformal prediction in regression: Conditionality, conjecture testing, and pertinent prediction intervals*. arXiv preprint arXiv:2109.12156.
- White, H., T. Kim, and S. Manganelli (2015). “VAR for VaR: Measuring tail dependence using multivariate regression quantiles”. In: *Journal of Econometrics* 187.3, pp. 169–188.
- Williams, W. H. and M. L. Goodman (1971). “A simple method for the construction of empirical confidence limits for economic forecasts”. In: *Journal of the American Statistical Association* 66.336, pp. 752–754.
- Xu, C. and Y. Xie (2021). “Conformal prediction interval for dynamic time-series”. In: *International Conference on Machine Learning*.
- (2023). “Sequential predictive conformal inference for time series”. In: *International Conference on Machine Learning*.
- Zaffran, M. et al. (2022). “Adaptive conformal predictions for time series”. In: *International Conference on Machine Learning*.

Table 1: Monte Carlo simulation: AR(1)

Standard Normal Errors $\epsilon_t \sim N(0, 1)$									
	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
CP1	0.036	0.019	0.014	0.076	0.060	0.054	0.124	0.110	0.103
CP2	0.021	0.015	0.012	0.063	0.053	0.051	0.109	0.103	0.101
FpB	0.017	0.014	0.012	0.057	0.053	0.052	0.104	0.105	0.102
FsB	0.015	0.013	0.012	0.054	0.052	0.051	0.102	0.103	0.102
PI-CP	0.957	0.971	0.977	0.877	0.895	0.898	0.784	0.797	0.800
PI-FpB	0.965	0.972	0.975	0.889	0.895	0.895	0.793	0.794	0.797
PI-FsB	0.970	0.974	0.976	0.894	0.897	0.897	0.797	0.797	0.798

Shifted χ^2 Errors $\epsilon_t \sim \chi^2(5)$									
	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
CP1	0.033	0.016	0.012	0.069	0.054	0.050	0.119	0.104	0.099
CP2	0.001	0	0	0.022	0.012	0.009	0.080	0.073	0.072
FpB	0.007	0.007	0.009	0.042	0.047	0.049	0.096	0.099	0.100
FsB	0.005	0.006	0.008	0.041	0.046	0.048	0.095	0.098	0.099
PI-CP	0.957	0.971	0.976	0.884	0.899	0.903	0.788	0.802	0.804
PI-FpB	0.974	0.978	0.978	0.900	0.900	0.899	0.797	0.799	0.799
PI-FsB	0.978	0.980	0.980	0.906	0.903	0.902	0.802	0.802	0.801

Note: Results from 1,000 replications of $y_t = 0.5y_{t-1} + \epsilon_t$ with $\epsilon_t \sim N(0, 1)$ (top panel) and ϵ_t drawn from a $\chi^2(5)$ distribution shifted such that $\mathbb{E}[\epsilon_t] = 0$ (bottom panel). The column header n denotes the sample size used for estimation (T in the text). Exceedance frequencies are computed from 1,000 draws of y_{n+1} for each replication and averaged across replications. CP1: one-sided conformal quantile (Algorithm 1); CP2: lower end of the two-sided conformal interval; FpB: non-studentized forward bootstrap with predictive residuals (Algorithm 3.2 in Pan and Politis (2016)); FsB: studentized forward bootstrap with fitted residuals (Algorithm 3.3 in Pan and Politis (2016)). The last three rows report the coverage of two-sided prediction intervals with nominal level $1 - 2\alpha$ obtained from two-sided conformal prediction and the two bootstraps.

Table 2: Monte Carlo simulation: AR(1) with exogenous regressors

	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
	$\kappa = 0.10$	$\kappa = 0.25$	$\kappa = 0.5$	$\kappa = 0.10$	$\kappa = 0.25$	$\kappa = 0.5$	$\kappa = 0.10$	$\kappa = 0.25$	$\kappa = 0.5$
CP1	0.013	0.011	0.008	0.043	0.042	0.041	0.105	0.084	0.073
CP2	0.010	0.008	0.006	0.040	0.038	0.038	0.099	0.076	0.066
FpB	0.008	0.007	0.001	0.036	0.031	0.025	0.102	0.076	0.047
FsB	0.010	0.012	0.010	0.048	0.050	0.053	0.112	0.100	0.084
PI-CP	0.981	0.987	0.988	0.910	0.923	0.928	0.810	0.838	0.859
PI-FpB	0.986	0.987	0.996	0.917	0.937	0.953	0.808	0.839	0.903
PI-FsB	0.983	0.979	0.982	0.896	0.899	0.899	0.793	0.797	0.824

Note: Results from 1,000 replications of $y_t = 0.5y_{t-1} + X'_{t-1}\beta + \epsilon_t$ with $\epsilon_t \sim N(0, 1)$, each element of the $n \times K$ regressor matrix X drawn from $N(0, 1)$ and β a vector of ones. The sample size is $n = 100$ (T in the text) and $\kappa = K/n$ measures the dimension of the problem. Exceedance frequencies are computed from 1,000 draws of y_{n+1} for each replication and averaged across replications. Models as in Table 1.

Table 3: Monte Carlo simulation: AR(1) with an outlier

	Standard Normal Errors $\epsilon_t \sim N(0, 1)$ with 1% outliers from $- N(\mu_0, 0.5) $								
	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
	$\mu_0 = 5$	$\mu_0 = 10$	$\mu_0 = 15$	$\mu_0 = 5$	$\mu_0 = 10$	$\mu_0 = 15$	$\mu_0 = 5$	$\mu_0 = 10$	$\mu_0 = 15$
CP1	0.009	0.009	0.009	0.051	0.051	0.049	0.095	0.097	0.097
CP2	0.009	0.007	0.008	0.045	0.040	0.035	0.086	0.080	0.072
FpB	0.006	0.006	0.006	0.045	0.043	0.039	0.090	0.090	0.087
FsB	0.005	0.005	0.005	0.041	0.031	0.022	0.086	0.078	0.064
PI-CP	0.981	0.983	0.980	0.902	0.905	0.908	0.808	0.808	0.807
PI-FpB	0.979	0.981	0.978	0.902	0.905	0.911	0.805	0.807	0.811
PI-FsB	0.982	0.988	0.987	0.908	0.925	0.941	0.812	0.829	0.847

Note: Results from 1,000 replications of $y_t = 0.5y_{t-1} + \epsilon_t$, where exactly one error in the sample is drawn from the outlier distribution $-|N(\mu_0, 0.5)|$ and the future errors are drawn from the mixture $(1 - \omega)N(0, 1) + \omega(-|N(\mu_0, 0.5)|)$ with $\omega = 0.01$. The sample size is $n = 100$ (T in the text). Exceedance frequencies are computed from 1,000 draws of y_{n+1} for each replication and averaged across replications. Models as in Table 1.

Table 4: Monte Carlo simulation: quantile loss for deciles relative to the DMQ

	Deciles								
	$\alpha=0.10$	$\alpha=0.20$	$\alpha=0.30$	$\alpha=0.40$	$\alpha=0.50$	$\alpha=0.60$	$\alpha=0.70$	$\alpha=0.80$	$\alpha=0.90$
n=50	1.055	1.033	1.038	1.050	1.051	1.053	1.054	1.036	1.046
n=100	1.014	1.009	1.015	1.010	1.014	1.027	1.025	1.024	1.024
n=200	1.060	1.030	1.019	1.020	1.030	1.015	0.995	0.999	0.996

Note: Quantile loss of CP1 (Algorithm 1 with an AR(4) point forecast) relative to the DMQ for the one-step-ahead forecast of each decile; values below one favor CP1. Data are generated by the DMQ of Catania and Luati (2023) with parameters, in their notation, $\phi = 0.95$, $\gamma = 0.10$, $\alpha = 0.01$ and $\beta = 0.7$. Samples of $n = 50, 100$ and 200 observations (T in the text) are used for estimation. Results are averages over 500 replications.

Table 5: Nowcasting US GDP: exceedance frequencies of 5% quantile nowcasts, odd weeks

	Weeks							
	1	3	5	7	9	11	13	15
BMF-SV	0.066	0.066	0.05	0.033	0.05	0.05	0.033	0.033
CP1-MIDAS	0.033	0.033	0.041	0.058	0.05	0.05	0.074	0.05
CP2-MIDAS	0.033	0.041	0.033	0.017 ***	0.017 ***	0.033	0.05	0.041
BOOT-MIDAS	0.017 ***	0.008 ***	0.017 ***	0.017 ***	0.017 ***	0.017 ***	0.008 ***	0.008 ***
DCP-BMF	0.099 *	0.099 *	0.124 **	0.091	0.091	0.099 *	0.099 *	0.099 *
FACI-MIDAS	0.025 *	0.025 *	0.033	0.041	0.041	0.025 *	0.041	0.025 *

Note: Empirical exceedance frequencies of real-time 5% quantile nowcasts for odd weeks, 1989Q4–2019Q3. BMF-SV: Bayesian mixed-frequency model with stochastic volatility; CP1-MIDAS: Algorithm 1 applied to the MIDAS nowcast; CP2-MIDAS: lower end of the two-sided conformal interval; BOOT-MIDAS: bootstrap of Aastveit, Foroni, and Ravazzolo (2017); DCP-BMF: one-sided distributional conformal prediction (Chernozhukov, Wüthrich, and Zhu 2021b) applied to the BMF; FACI-MIDAS: FACI (Gibbs and Candès 2024) wrapped around CP1-MIDAS. *, ** and *** denote rejection of the null hypothesis of correct coverage at the 10%, 5% and 1% level, based on a two-sided t -test.

Table 6: Nowcasting US GDP: quantile scores of 5% quantile nowcasts, odd weeks

	Weeks							
	1	3	5	7	9	11	13	15
BMF-SV	0.252	0.244	0.176	0.162	0.19	0.174	0.192	0.205
CP1-MIDAS	0.908	0.853	1.331	1.136	0.916	1.053	0.846	0.901
CP2-MIDAS	0.893	0.842	1.324 **	1.325 ***	0.966	1.043	0.826	0.892
BOOT-MIDAS	1.080	0.982	1.452 ***	1.494 ***	1.285 ***	1.362 ***	1.325 ***	1.162 **
DCP-BMF	1.209	1.242	1.42	1.327	1.114	1.135	1.047	0.948
FACI-MIDAS	0.938	0.934	1.495 **	1.216 *	0.942	1.085	0.806	0.922

Note: Average quantile score (4) of real-time 5% quantile nowcasts for odd weeks, 1989Q4–2019Q3. The first row reports the average score of the BMF-SV; all other rows report ratios relative to the BMF-SV, so that values below one indicate a lower loss than the BMF-SV. Models as in Table 5. *, ** and *** denote significance at the 10%, 5% and 1% level of a Diebold–Mariano test of equal quantile scores relative to the BMF-SV.

Table 7: Microsoft returns: aggregate quantile loss relative to the DMQ

	Aggregate Percentiles									
	[0.01,0.10]	[0.10,0.20]	[0.20,0.30]	[0.30,0.40]	[0.40,0.50]	[0.50,0.60]	[0.60,0.70]	[0.70,0.80]	[0.80,0.90]	[0.90,0.99]
FACI	0.987	0.994	0.995	0.994	0.995	0.992	0.987	0.980	0.971	0.953
CP1	1.003	0.997	0.998	0.994	0.996	0.995	0.992	0.987	0.984	0.973

Note: Aggregate quantile loss of CP1 (Algorithm 1 with an AR(4) point forecast) and of FACI wrapped around CP1, relative to the DMQ, so that values below one indicate a lower loss than the DMQ. The aggregate loss of each decile is the sum of the quantile losses of the percentiles in that decile. FACI belongs to the 75% model confidence set of Hansen, Lunde, and Nason (2011) for all deciles; CP1 and the DMQ belong to it for the first five deciles only.

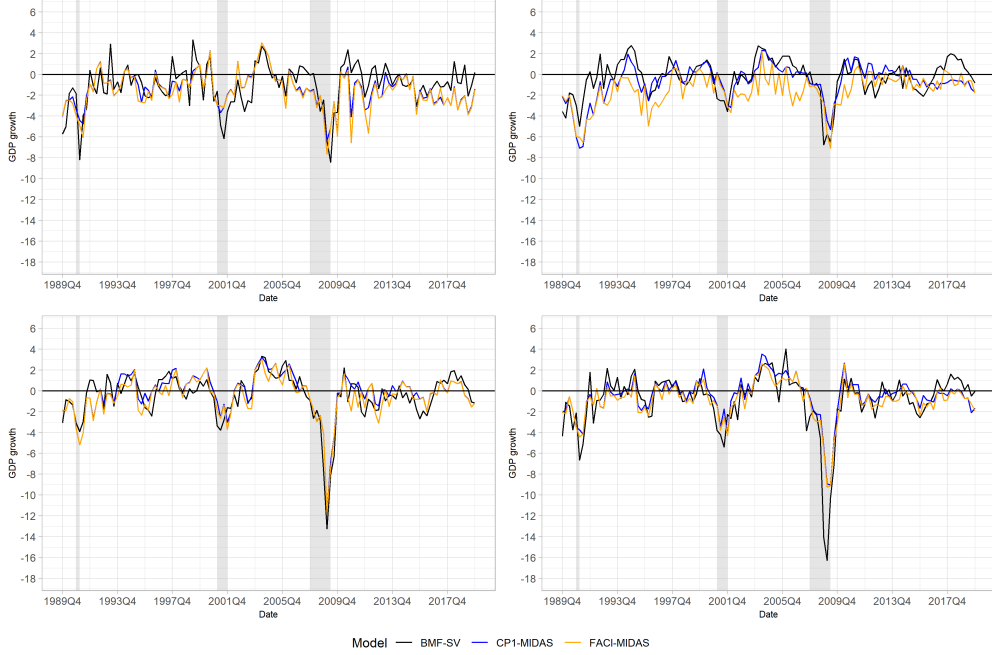


Figure 1: Real-time 5% quantile nowcasts of US GDP growth, 1989Q4–2019Q3

Note: Real-time out-of-sample 5% quantile nowcasts of GDP growth from the BMF-SV, CP1-MIDAS and FACI-MIDAS. Shaded areas correspond to NBER-dated recessions. From left to right and top to bottom: weeks 1, 5, 10 and 15.

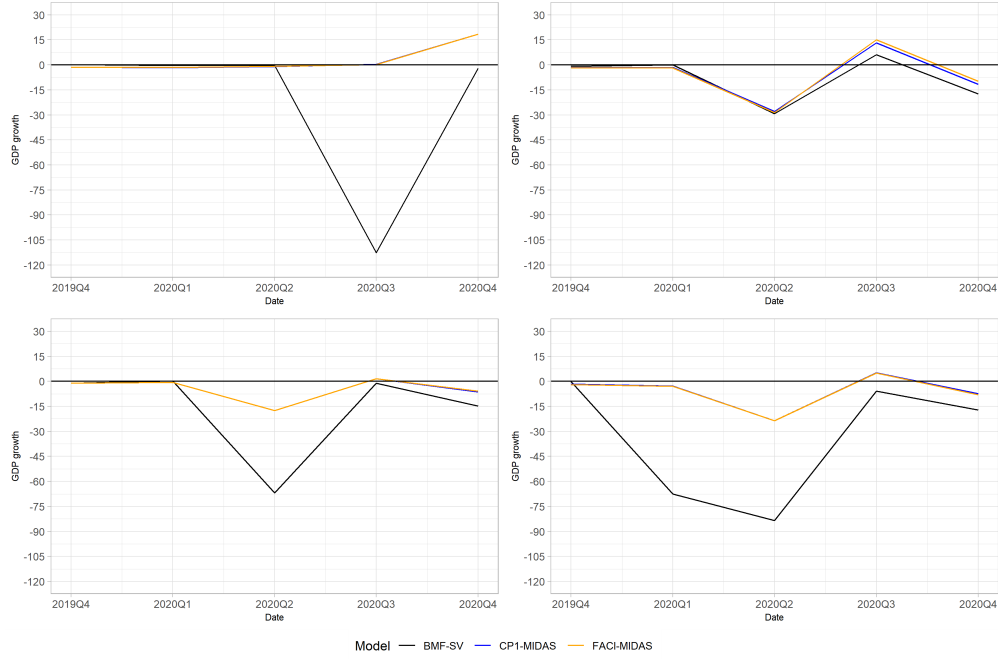


Figure 2: Real-time 5% quantile nowcasts of US GDP growth during COVID-19, 2019Q4–2020Q4

Note: Real-time out-of-sample 5% quantile nowcasts of GDP growth from the BMF-SV, CP1-MIDAS and FACI-MIDAS. Shaded areas correspond to NBER-dated recessions. From left to right and top to bottom: weeks 1, 5, 10 and 15.

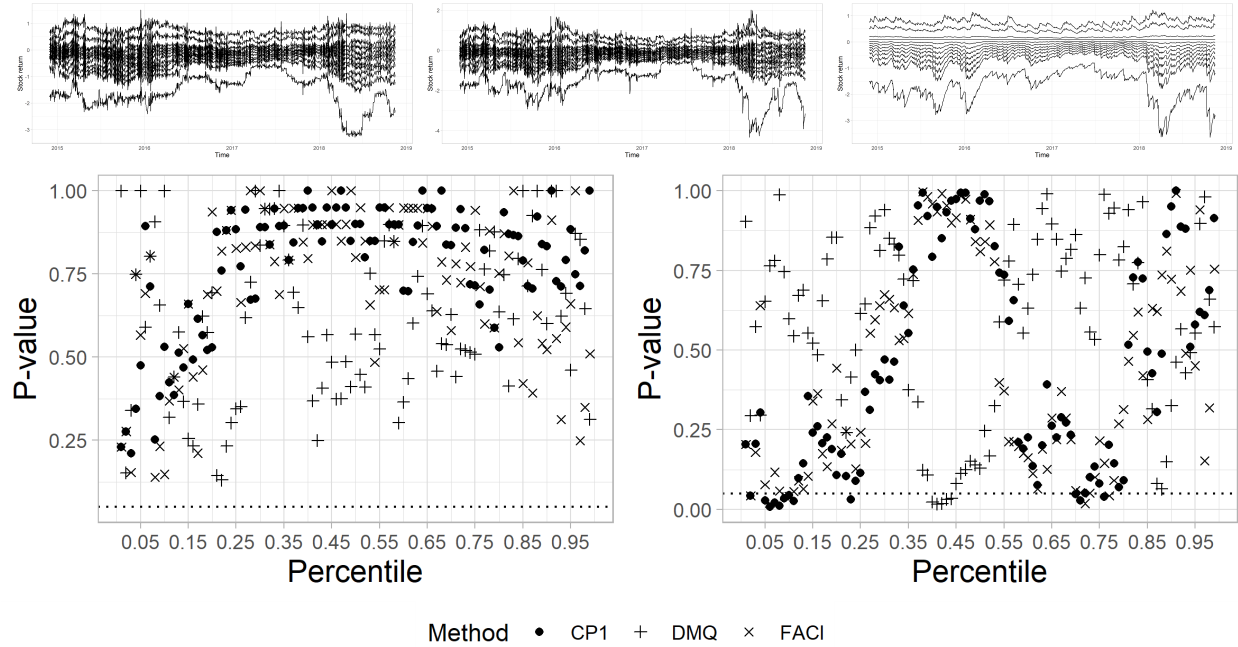


Figure 3: Microsoft returns: quantile forecasts and backtests

Note: Top panel: one-step-ahead out-of-sample forecasts of the 5%, 10%, ..., 95% quantiles of daily Microsoft returns from CP1 (Algorithm 1 with an AR(4) point forecast), FACI wrapped around CP1 and the DMQ, from left to right. Bottom panel: p-values of the unconditional and conditional backtests of Kupiec (1995) and Christoffersen (1998) for each percentile.

A Proof of Proposition 1

Part (i). Let R denote the rank of e_t among the $n + 1$ errors in $\mathcal{E}_t \cup \{e_t\}$, so that $R = 1 + |\{s \in S_t : e_s < e_t\}|$, where S_t indexes the errors in $\mathcal{E}_t$. Because the errors are exchangeable and almost surely distinct, every ordering of the $n + 1$ errors is equally likely, and R is uniformly distributed on $\{1, \dots, n + 1\}$ (see, e.g., Vovk, Gammernan, and Shafer 2005; Kuchibhotla 2021). Since $y_t - \hat{y}_t = e_t$,

$$y_t \leq \hat{y}_t + e_{(k)} \iff e_t \leq e_{(k)} \iff |\{s \in S_t : e_s < e_t\}| \leq k - 1 \iff R \leq k,$$

where the second equivalence uses that the errors are distinct: $e_t \leq e_{(k)}$ holds if and only if fewer than k elements of $\mathcal{E}_t$ are smaller than e_t . Hence $P(y_t \leq \hat{y}_t + e_{(k)}) = P(R \leq k) = k/(n + 1)$.

Part (ii). For the conservative version, $\hat{q}_t(\alpha) = \hat{y}_t + e_{(k)}$ with $k = \lfloor \alpha(n + 1) \rfloor$, which satisfies $\alpha(n + 1) - 1 < k \leq \alpha(n + 1)$ and $k \geq 1$ whenever $\alpha \geq 1/(n + 1)$. The result follows from part (i).

Part (iii). For the interpolated version, $\hat{Q}_\alpha(\mathcal{E}_t)$ lies between $e_{(j)}$ and $e_{(j+1)}$, where $j = \lfloor (n - 1)\alpha \rfloor + 1 \leq n - 1$ for $\alpha < 1$. Hence

$$\{y_t \leq \hat{y}_t + e_{(j)}\} \subseteq \{y_t \leq \hat{q}_t(\alpha)\} \subseteq \{y_t \leq \hat{y}_t + e_{(j+1)}\},$$

and part (i) yields the bounds $j/(n + 1)$ and $(j + 1)/(n + 1)$. Since $(n - 1)\alpha < j \leq (n - 1)\alpha + 1$, the lower bound is at least $\alpha - 2\alpha/(n + 1)$ and the upper bound at most $\alpha + 2(1 - \alpha)/(n + 1)$, which implies $|P(y_t \leq \hat{q}_t(\alpha)) - \alpha| \leq 2/(n + 1)$.

Part (iv). Both versions of $\hat{Q}_\alpha(\mathcal{E})$ are non-decreasing in α for any fixed set $\mathcal{E}$, and all quantile levels share the same point nowcast $\hat{y}_t$ and the same set $\mathcal{E}_t$. $\square$

B Additional results

B.1 Nowcasting results for even weeks

Table 8: Nowcasting US GDP: exceedance frequencies of 5% quantile nowcasts, even weeks

	Weeks						
	2	4	6	8	10	12	14
BMF-SV	0.058	0.066	0.041	0.041	0.066	0.033	0.05
CP1-MIDAS	0.025 *	0.033	0.066	0.066	0.041	0.058	0.058
CP2-MIDAS	0.033	0.033	0.025 *	0.017 ***	0.017 ***	0.033	0.05
BOOT-MIDAS	0.017 ***	0.017 ***	0.017 ***	0.017 ***	0.008 ***	0.008 ***	0.008 ***
DCP-BMF	0.091	0.083	0.107 **	0.099 *	0.099 *	0.099 *	0.083
C-MIDAS	0.033	0.033	0.041	0.033	0.025 *	0.033	0.05

Note: As Table 5 for even weeks, 1989Q4–2019Q3. The row labelled C-MIDAS reports FACI-MIDAS.

Table 9: Nowcasting US GDP: quantile scores of 5% quantile nowcasts, even weeks

	Weeks						
	2	4	6	8	10	12	14
BMF-SV	0.254	0.225	0.17	0.174	0.188	0.167	0.204
CP1-MIDAS	0.857	0.881	1.269	1.078	0.896	1.064	0.876
CP2-MIDAS	0.808 **	0.882	1.39 ***	1.222 ***	0.929	1.057	0.872
BOOT-MIDAS	1.014	1.066	1.461 ***	1.384 ***	1.25 ***	1.388 ***	1.164 **
DCP-BMF	1.133	1.155	1.392	1.249	1.086	1.235	0.995
C-MIDAS	0.939	0.915	1.38 *	1.207 *	0.915	1.061	0.89

Note: As Table 6 for even weeks, 1989Q4–2019Q3. The row labelled C-MIDAS reports FACI-MIDAS.

B.2 Alternative bootstrap procedures for the nowcasting application

Table 10 reports the exceedance frequencies of the two forward bootstraps of Section 3 applied to the MIDAS model, together with BOOT-MIDAS. All three bootstraps overcover in every week.

Table 10: Nowcasting US GDP: exceedance frequencies of alternative bootstrap procedures

	Weeks							
	1	3	5	7	9	11	13	15
BOOT-MIDAS	0.017 ***	0.008 ***	0.017 ***	0.017 ***	0.017 ***	0.017 ***	0.008 ***	0.008 ***
FpB	0.017 ***	0.008 ***	0.017 ***	0 ***	0 ***	0 ***	0 ***	0.008 ***
FsB	0.017 ***	0.017 ***	0.025 *	0.008 ***	0.017 ***	0.017 ***	0.017 ***	0.008 ***

	Weeks						
	2	4	6	8	10	12	14
BOOT-MIDAS	0.017 ***	0.017 ***	0.017 ***	0.017 ***	0.008 ***	0.008 ***	0.008 ***
FpB	0.008 ***	0 ***	0 ***	0 ***	0 ***	0 ***	0 ***
FsB	0.017 ***	0.017 ***	0.025 *	0.008 ***	0.017 ***	0.017 ***	0.017 ***

Note: Empirical exceedance frequencies of real-time 5% quantile nowcasts for all weeks, 1989Q4–2019Q3. BOOT-MIDAS: bootstrap of Aastveit, Foroni, and Ravazzolo (2017); FpB: non-studentized forward bootstrap with predictive residuals (Algorithm 3.2 in Pan and Politis (2016)); FsB: studentized forward bootstrap with fitted residuals (Algorithm 3.3 in Pan and Politis (2016)). *, ** and *** as in Table 5.

B.3 Window length in the multiple quantile application

Table 11 and Figure 4 report the results of Section 5 for rolling windows of $w = 50, 100$ and 200 past errors. For $w < 100$, some forecast percentiles coincide by construction (Remark 3). Two features stand out. First, the aggregate decile loss decreases as the window shrinks. This is mostly due to the aggregation into deciles and does not hold for the individual percentile losses, which are not reported to keep the table readable. Second, there is a trade-off between unconditional and conditional backtesting performance. Longer windows perform better in the unconditional test, shorter windows in the conditional test, because a short window lets the forecast quantiles adapt to local changes in the error distribution. In the main text we use the intermediate value $w = 100$, which balances a low quantile loss against good unconditional and conditional backtesting properties.

Table 11: Microsoft returns: aggregate quantile loss for different window lengths

	Aggregate Percentiles									
	[0.01,0.10)	[0.10,0.20)	[0.20,0.30)	[0.30,0.40)	[0.40,0.50)	[0.50,0.60)	[0.60,0.70)	[0.70,0.80)	[0.80,0.90)	[0.90,0.99)
w=50	0.961	0.987	0.989	0.993	0.997	0.997	0.993	0.984	0.975	0.954
w=100	1.003	0.997	0.998	0.994	0.996	0.995	0.992	0.987	0.984	0.973
w=200	1.016	1.009	1.009	1.004	1.002	1.001	0.998	0.995	0.994	1.001

Note: As Table 7 for CP1 with rolling windows of $w = 50, 100$ and 200 past errors. Percentile losses are aggregated into decile losses.

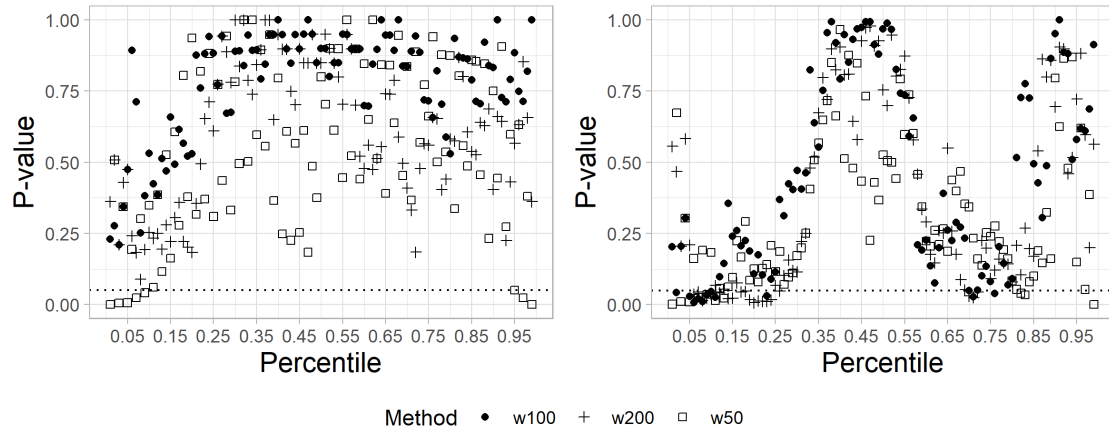


Figure 4: Microsoft returns: backtests for different window lengths

Note: p-values of the unconditional and conditional backtests of Kupiec (1995) and Christoffersen (1998) for CP1 with rolling windows of different lengths.