Received: 30 August 2019 | Revised: 8 December 2020 | Accepted: 10 March 2021

DOI: 10.1002/jae.2833

RESEARCH ARTICLE

WILEY Journal of APPLIED ECONOMETRICS

# Interpretation of point forecasts with unknown directive

**Patrick Schmidt**[1] | **Matthias Katzfuss**[2] | **Tilmann Gneiting**[3,4]

[1]Department of Psychology, University of Zurich, Zürich, Switzerland

[2]Department of Statistics, Texas A&M University, College Station, Texas, USA

[3]Computational Statistics, Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

[4]Institute for Stochastics, Karlsruhe Institute of Technology, Karlsruhe, Germany

**Correspondence**
Patrick Schmidt, Department of Psychology, University of Zurich, Rämistrasse 71, 8006 Zürich, Switzerland.
Email: patrickwolfgang.schmidt@uzh.ch

**Funding information**
Klaus Tschira Foundation; European Union Seventh Framework Programme, Grant/Award Number: 290976; US National Science Foundation (NSF), Grant/Award Number: DMS–1521676; NSF CAREER, Grant/Award Number: DMS–1654083

**Abstract**

Point forecasts can be interpreted as functionals (i.e., point summaries) of predictive distributions. We extend methodology for the identification of the functional based on time series of point forecasts and associated realizations. Focusing on state-dependent quantiles and expectiles, we provide a generalized method of moments estimator for the functional, along with tests of optimality under general joint hypotheses of functional relationships and information bases. Our tests are more flexible, and in simulations better calibrated and more powerful than existing solutions. In empirical examples, economic growth forecasts and model output for precipitation are indicative of overstatement in anticipation of extreme events.

**KEYWORDS**

expectile, identifying moment conditions, information set, loss function, optimality of point forecasts, quantile

## 1 | INTRODUCTION

Forecasts are frequently the basis of crucial decisions. Yet they are fraught with uncertainty due to imperfections in the observation, understanding, and modeling of the underlying mechanisms. To account for this uncertainty, it is increasingly being recognized that forecasts ought to be probabilistic in nature (Gneiting & Katzfuss, 2014). If forecasts are issued in the form of full predictive distributions, there are well-established methods for computing the Bayes act in any given decision problem, for testing optimality, and for comparing and ranking competing forecasting methods.

However, single-valued point forecasts remain ubiquitous. Their interpretation requires assumptions on the decision process or directive that the forecasters used in order to generate the point predictions (Elliott et al., 2005; Elliott & Timmermann, 2008; Engelberg et al., 2009; Manski, 2016). A directive can be expressed through a functional (i.e., a real-valued summary) of the predictive distribution, and it is a widely used assumption that the reported functional is the mean value or expectation. However, there is often little justification for this supposition. Knowledge of the functional used to generate the forecast is important, as it allows for proper interpretation, evaluation, testing, and comparison of point predictions (Gneiting, 2011).

Here, we address the setting of point forecasts with unknown directive, for which the forecaster implicitly (only) reports a certain functional of the predictive distribution. This scenario can arise with expert forecasts or response items in

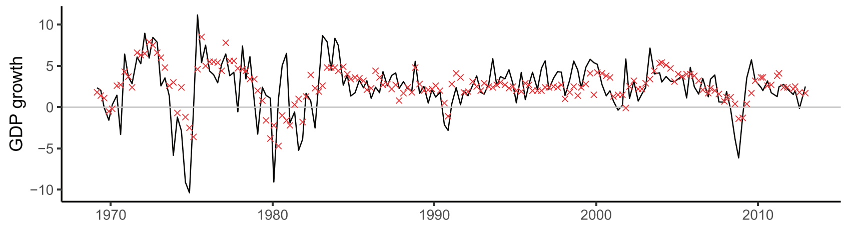


**FIGURE 1** Time series of one-quarter-ahead Greenbook forecasts (crosses) and respective observations (solid line) of real gross domestic product (GDP) growth in the United States (in percent) [Colour figure can be viewed at wileyonlinelibrary.com]

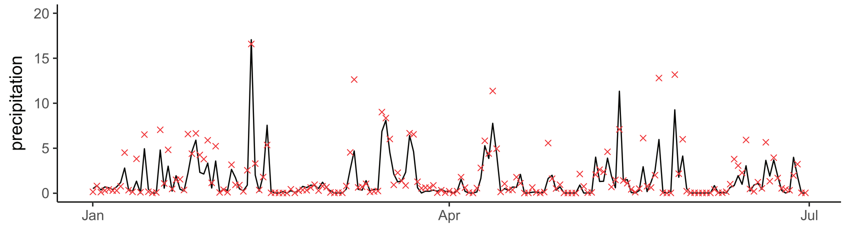


**FIGURE 2** Time series of 24-h-ahead European Centre for Medium-Range Weather Forecasts (ECMWF) forecasts (crosses) and respective observations (solid line) of daily accumulated precipitation (in millimeter) at London, UK, in 2013 [Colour figure can be viewed at wileyonlinelibrary.com]

surveys. Another important example is output from complex computer models, such as in weather and climate prediction, which are often tuned by multiple individuals to achieve forecasts that model developers or decision makers perceive as optimal, in ways that might neither be transparent nor explicitly defined. Such forecasts would be most informative if the user knew the directive under which the forecast was issued. Our goal here is to estimate the functional from time series data of point forecasts and associated realizations and to construct tests regarding the properties of the functional. The type of data encountered in this setting is illustrated in Figure 1, which displays one-quarter-ahead Greenbook gross domestic product (GDP) forecasts of the US Federal Reserve, and in Figure 2, which shows 24-h-ahead forecasts of daily accumulated precipitation at London, UK, from the high-resolution run operated by the European Centre for Medium-Range Weather Forecasts (ECMWF). While we focus on time series data, adaptations to cross-sectional data are straightforward. Once the functional has been estimated, the point forecasts can be coherently interpreted and be compared with other point or probability forecasts, and constructive feedback can be given to model developers.

We build on extant work that focuses on estimation of the loss function underlying a forecast. The pioneering work of Elliott et al. (2005) provides a generalized method of moments (GMM) estimator for piecewise linear and quadratic loss functions. Patton and Timmermann (2007) apply methods of this type to the US Federal Reserve's Greenbook GDP forecasts with a new class of loss functions, which consists of quadratic splines that depend on a state variable in flexible ways. Recently, asymmetric piecewise linear and piecewise quadratic loss functions have been estimated in various economic applications (Christodoulakis & Mamatzakis, 2008; Capistrán, 2008; Elliott et al., 2008; Fritsche et al., 2015; Guler et al., 2017; Krol, 2013; Lieli et al., 2019; Pierdzioch et al., 2013; Wang & Lee, 2014).

Here, we argue that the loss function is not identifiable from point forecasts and realizations only. Specifically, if a functional is defined via a loss function, then there exists a whole class of nontrivially distinct loss functions that define the very same functional (Ehm et al., 2016; Steinwart et al., 2014). It is therefore futile to identify the shape of the loss, given that all these functions lead to the same functional-forecast and moment conditions. Hence, we formalize notions of optimality for point forecasts in terms of functionals rather than loss functions. We consider single-valued scalar functionals throughout, although the results extend to set-valued and multivariate cases under additional technical considerations. For estimation, we focus on parametric models for quantiles and expectiles, where the level of asymmetry depends on state variables. We propose a GMM estimator and apply standard GMM theory to show consistency and asymptotic normality under mild assumptions. We also discuss testing of forecast optimality and other forecast properties. The approach of Elliott et al. (2005) can be viewed as a special case of our method that assumes quantile or expectile forecasts at a constant level. Our state-dependent functionals allow the consistent interpretation of a much wider range of point forecasts. In comparison to Patton and Timmermann (2007), our methods yield improved interpretability and nominally sized tests.

Our approach is based on the insight that one can only hope to identify the functional relationship between a forecast and the data-generating distribution. Such a reduced-form approach avoids a priori hypotheses about the underlying reasons, which can then be investigated subsequently, once a quantitative relationship has been established. Forecast optimality needs to be defined and studied relative to information sets (Holzmann & Eulert, 2014). In our approach, tests operate under general hypotheses of forecast optimality relative to information sets, and the choice of instrumental variables for the GMM estimator affects and determines the joint conditions tested for. The addition of information into the instrument vector yields stronger hypotheses of forecast optimality and increases power against suboptimal forecasts. In real data and simulation examples, we elucidate the role of the information set, and we illustrate that the proposed optimality test can be used to distinguish information bases. Importantly, while our functional-based approach provides a complementary perspective that is both convenient and flexible, it does not invalidate approaches based on loss functions.

The remainder of the paper is organized as follows. In Section 2, we review notions of optimality for general types of point forecasts and discuss the identifying moment conditions upon which our approach is based. For the remainder of the paper, we focus on the particular case of possibly state-dependent quantiles and expectiles. In Section 3, we introduce a parametric GMM estimator in the time series setting, study its large sample behavior, and discuss the aforementioned tests of forecast optimality. The Monte Carlo studies in Section 4 serve to compare our approach to existing solutions. Section 5 is dedicated to empirical studies, where our approach yields accessible and scientifically relevant insights. In particular, we posit that Greenbook GDP forecasts of the US Federal Reserve and ECMWF forecasts of daily accumulated precipitation can be interpreted as state-dependent quantiles or expectiles, and we observe a pronounced tendency to exaggerate in anticipation of tail events. The paper closes with a discussion in Section 6. The Supporting Information contains technical results and additional detail in Sections S1 through S9.

## 2 | OPTIMALITY OF POINT FORECASTS AND IDENTIFYING MOMENT CONDITIONS

Consider a real-valued random variable $Y$ and a corresponding point forecast $X$, which is based on the information available to the forecaster, as encoded by some $\sigma$-algebra $\mathcal{F}$. Commonly, a point forecast is interpreted as the mean of the conditional distribution $\mathcal{L}(Y|\mathcal{F})$, that is,

$$X = \mathbb{E}[Y|\mathcal{F}].$$

Here and throughout the paper, equality of random variables is understood to hold almost surely. Proceeding to a more general framework, let $\alpha : \mathcal{P} \mapsto \mathbb{R}$ be a functional, that is, a single-valued mapping from some class $\mathcal{P}$ of probability distributions to the real line (Horowitz & Manski, 2006; Huber & Ronchetti, 2009, p. 9). A functional $\alpha$ is symmetric if, for every symmetric distribution $P \in \mathcal{P}$ with symmetry point $c$, it holds that $c = \alpha(P)$. Prominent alternatives to the mean functional are symmetric functionals, like the median, or asymmetric generalizations, such as quantiles, expectiles, and generalized quantiles (Bellini et al., 2014). Throughout, we use the short notation $\alpha(Y \mid \mathcal{F})$ for $\alpha(\mathcal{L}(Y \mid \mathcal{F}))$.

**Definition 1** (optimal $\alpha$-forecast)**.** A random variable $X$ is an optimal $\alpha$-forecast of $Y$ relative to the information set $\mathcal{F}$ if

$$X = \alpha(Y \mid \mathcal{F}).$$

Importantly, an optimal point forecast satisfies a condition in terms of both the functional $\alpha$ *and* the information set $\mathcal{F}$, in that the functional is applied to the conditional distribution $\mathcal{L}(Y \mid \mathcal{F})$. Hence, optimal $\alpha$-forecasts may differ from each other if they are optimal with respect to distinct information sets.

Crucially, we consider the situation in which the functional used by the forecaster *and* the conditional distribution $\mathcal{L}(Y \mid \mathcal{F})$ is unknown. In line with seminal work on professional economic forecasters (Elliott et al., 2005; Patton & Timmermann, 2007), we merely assume that the unknown conditional distribution constitutes a predictive distribution consistent with some information set $\mathcal{F}$.

Next, we introduce notation. If $R$ is a random variable or random vector and $Q$ is an integrable random variable, the relation $R \in \mathcal{F}_Q$ indicates that $R$ is $\mathcal{F}$-measurable and both $R$ and $QR$ are (componentwise) integrable. The relation $R \in \mathcal{F}$ means that $R$ is $\mathcal{F}$-measurable and integrable. As usual, we write $\sigma(W)$ for the information set generated by the random vector $W$. Uppercase letters represent random variables and corresponding lowercase letters their realizations. Finally, the partial derivative of a function $g(x, y)$ with respect to $x$ is denoted $g_{(x)}(x, y)$.

Standard properties of conditional expectations (e.g., Billingsley, 1995, section 34) yield identifying moment conditions for an optimal mean forecast relative to the information set $\mathcal{F}$, namely,

$$X = \mathbb{E}[Y|\mathcal{F}] \iff \mathbb{E}[(X - Y)W] = 0 \quad \text{for all} \quad W \in \mathcal{F}_{X-Y},$$

where the components of the random vector $W$ (henceforth called *instruments*) represent information $\mathcal{F}$ available to the forecaster when the prediction is issued. This property of optimal mean forecasts generalizes to optimal $\alpha$-forecasts.

Specifically, for every sufficiently regular functional $\alpha : \mathcal{P} \mapsto \mathbb{R}$, there exists a function $V_\alpha$ identifying the optimal $\alpha$-forecast, that is,

$$X = \alpha(Y|\mathcal{F}) \iff \mathbb{E}[V_\alpha(X, Y)W] = 0 \text{ for all } W \in \mathcal{F}_{V_\alpha(X,Y)}. \tag{1}$$

We refer to (1) as the *identifying moment conditions* for an optimal $\alpha$-forecast and give rigorous versions thereof in Section S1. The identification function $V_\alpha$ is unique up to an $\mathcal{F}$-measurable multiplicative factor (Steinwart et al., 2014, theorem 8), so the set of the arising moment conditions does not depend on its choice. The moment conditions are the foundation of the methodology developed hereinafter. In particular, they allow for tests of whether a point forecast is an optimal $\alpha$-forecast relative to a specific information set. In the hypothetical limit of an infinite supply of data and instruments, a nonrejection of the test is a sufficient condition for optimality.

In contrast to our approach, extant work (e.g., Elliott et al., 2005) has typically defined optimal point forecasts via a loss function $L$, namely, as

$$X = \arg\min_{x \in \mathbb{R}} \mathbb{E}[L(x, Y) \mid \mathcal{F}].$$

Under regularity conditions, this specifies a well-defined optimal $\alpha_L$-forecast, where the functional $\alpha_L$ is defined as

$$\alpha_L : \mathcal{P} \mapsto \mathbb{R}, \; P \mapsto \arg\min_{x \in \mathbb{R}} \mathbb{E}_{Y \sim P}[L(x, Y)],$$

for a suitable class $\mathcal{P}$ of probability distributions. For example, the mean functional can be defined as the minimizer of expected quadratic loss, $L(x, y) = (x - y)^2$, for probability distributions with finite second moments. While some functionals, such as the expected shortfall and the mode, do not admit definitions via loss functions relative to broad classes of probability distributions (Gneiting, 2011; Heinrich, 2014), every functional $\alpha$ with identifying moment conditions (1) can be defined via a loss function $L$ under weak conditions, and the identification function $V_\alpha$ derives from the partial derivative $L_{(x)}(x, y)$ (Fissler & Ziegel, 2016, theorem 3.2; Steinwart et al., 2014, theorem 8). In this setting, the literature commonly states necessary (but not sufficient) conditions for optimality (e.g., Diebold & Lopez, 1996; Patton & Timmermann, 2007, 2010). As a notable exception, proposition 1 of Elliott et al. (2005) supplies identifying moment conditions of the general form (1) that apply to linear predictors under piecewise linear and piecewise quadratic loss.

As noted, our rationale for the shift of emphasis from loss functions to functionals is an identifiability problem. For example, while the mean functional minimizes expected quadratic loss, given any convex and differentiable function $\varphi$ the loss function $L(x, y) = \varphi(y) - \varphi(x) - \varphi'(x)(y - x)$ also induces an optimal mean forecast (Savage, 1971). In this light, a more compelling approach is to estimate functionals, as they are identified to the extent that they differ on the class of the arising conditional distributions (Lieli & Stinchcombe, 2013). For instance, the mean and the median are distinct functionals in general, but if all considered distributions are symmetric, the two functionals are identical and cannot be identified. In what follows, we focus on optimal forecasts in the form of either quantiles or expectiles, to allow for unique identification without imposing unduly strong assumptions on the data-generating process.

## 3 | PARAMETRIC ESTIMATION AND TESTING OF STATE-DEPENDENT QUANTILES AND EXPECTILES

We turn to parametric estimation of possibly time-varying functionals. Consider a stochastic process $(X_t, Y_t, Z_t)_{t \in \mathbb{Z}}$ of forecasts, observations, and covariates, for which we have a sample path $(x_t, y_t, z_t)_{t=1,\ldots,T}$. Our goal is to infer the functional represented by the point forecasts.

We assume that at each point in time an optimal forecast is issued, that is, $X_t = \alpha_t(Y_t \mid \mathcal{F}_t)$ for $t \in \mathbb{Z}$. In the situation of an $h$-step-ahead forecast, the available information is typically generated by lagged variables of the outcome and the vector-valued covariate, so that $\mathcal{F}_t = \sigma(\{Y_u, Z_u : u \leq t - h\})$. For ease of notation, statements about all time points are often denoted without subscripts. For example, we write $X = \alpha(Y \mid \mathcal{F})$ instead of $X_t = \alpha_t(Y_t \mid \mathcal{F}_t)$ for $t \in \mathbb{Z}$.

Extending Definition 1, we allow the functional $\alpha$ to depend on the current situation, as represented by some $\mathcal{F}_t$-measurable state variable $S_t$, for a *state-dependent* functional. For example, in the aforementioned situation of an $h$-step-ahead forecast, $S_t$ might include the most recent observation, $Y_{t-h}$, components of the covariate vector, $Z_{t-h}$, or the current forecast, $X_t$. Asymmetric and state-dependent point forecasts can arise for a variety of reasons, including varying preferences of the forecaster, asymmetric information, and nonlinear transformations of the data, and we refer to

Section S2 for details. In the following, we assume that the true functional is a state-dependent quantile or state-dependent expectile of the conditional distribution $\mathcal{L}(Y \mid \mathcal{F})$. As elaborated in Section S3, this assumption is of surprising flexibility and generality.

## 3.1 | State-dependent quantiles and expectiles

We first discuss identifying moment conditions for constant quantile and expectile forecasts, which correspond to first order conditions derived in Elliott et al. (2005, pp. 1100–1111) under the loss function approach. The $\tau$-quantile functional $q_\tau(P)$ of a distribution $P$ with continuous and strictly increasing cumulative distribution function is the unique solution $x$ to the equation $P((-\infty, x]) = \tau \in (0, 1)$. In our setting, we can express this directly in terms of the identification function of the $\tau$-quantile, namely, $V_\tau(x, y) = \mathbb{1}(y \le x) - \tau$:

$$X = q_\tau(Y|\mathcal{F}) \iff \mathbb{E}[(\mathbb{1}(Y \le X) - \tau)W] = 0 \quad \text{for all} \quad W \in \mathcal{F}.$$

When $\tau = 1/2$, we recover the classical result that a median forecast is optimal if, and only if, the sign of the forecast error is uncorrelated with any random variable in the forecaster's information set. For technical details, see Section S1.

While quantiles are asymmetric generalizations of the median, expectiles are asymmetric generalizations of the mean. Specifically, Newey and Powell (1987) introduced the $\tau$-expectile $e_\tau(P)$ of a nondegenerate distribution $P$ with finite mean as the unique solution $x$ to the equation

$$\frac{\tau}{1-\tau} = \frac{\int_{-\infty}^{x}(x-y)\mathrm{d}P(y)}{\int_{x}^{\infty}(y-x)\mathrm{d}P(y)},$$

where $\tau \in (0, 1)$. In our setting, this is equivalent to

$$X = e_\tau(Y|\mathcal{F}) \iff \mathbb{E}[|\mathbb{1}(Y \le X) - \tau|(Y - X)W] = 0 \quad \text{for all} \quad W \in \mathcal{F}_{X-Y},$$

which reveals the corresponding identification function $V_\tau(x, y) = |\mathbb{1}(y \le x) - \tau|(y - x)$. When $\tau = 1/2$, we see that a mean forecast is optimal if, and only if, the forecast error is uncorrelated with any random variable in the forecaster's information set.

We allow for additional flexibility and let the level $\tau$ of the quantile or expectile functional depend on the state $s$ via a parametric function $m(s, \theta)$.

**Definition 2** (specification model)**.** Let $\Theta$ be a subset of $\mathbb{R}^p$ and suppose that the state variable $s$ takes values in $\mathbb{R}^k$. A specification model is a measurable function $m(s, \theta)$ that maps $\mathbb{R}^k \times \Theta$ into the unit interval $(0, 1)$.

Our key assumption then is that $X_t$ is an optimal quantile or expectile forecast with state-dependent level $m(S_t, \theta_*)$ for some $\theta_* \in \Theta$, that is, in the case of quantiles

$$X_t = q_{m(S_t,\theta_*)}(Y_t \mid \mathcal{F}_t), \tag{2}$$

for $t \in \mathbb{Z}$, and analogously for expectiles. Crucially, we assume that the state variable $S_t$ is known to the forecaster at the time when the forecast is issued.

We say that a specification model $m(s, \theta)$ is continuous (continuously differentiable) if it is continuous (continuously differentiable) in $\theta \in \Theta$ for every $s \in \mathbb{R}^k$. Examples are given in Table 1, where the state $s$ is assumed to be real valued. The *constant* model assumes that the forecaster always states the $\theta$-quantile or $\theta$-expectile, where $\theta \in (0, 1)$ is unknown and needs to be estimated. This model has been implemented under the loss function paradigm in much previous work (Christodoulakis & Mamatzakis, 2008; Elliott et al., 2005; Fritsche et al., 2015; Krol, 2013; Pierdzioch et al., 2013). The related *fixed* model assumes that the forecaster's directive is known to be the quantile or expectile at the true level $\theta_*$. If $\theta_* = 1/2$, this is equivalent to assuming a median or mean forecast.

**TABLE 1** Specification models for a real-valued state variable $s$

| Name | Model | Parameter space $\Theta$ |
|---|---|---|
| Constant | $m(s,\theta) = \theta$ | $(0,1)$ |
| Break | $m(s,\theta) = \Phi(\mathbb{1}(s \le \theta_2)\theta_0 + \mathbb{1}(s > \theta_2)\theta_1)$ | $\mathbb{R}^3$ |
| Linear | $m(s,\theta) = \Phi(\theta_0 + \theta_1 s)$ | $\mathbb{R}^2$ |
| Periodic | $m(s,\theta) = \Phi(\theta_0 + \theta_1 \sin(2\pi s/\theta_2))$ | $\mathbb{R}^2 \times (0,\infty)$ |

For the state-dependent models, we employ the probit link, where the CDF $\Phi$ of the standard normal distribution ensures a quantile or expectile level in the unit interval. Alternatively, the logit link or a related formulation could be used. Throughout, we consider specification models that nest the constant model and, therefore, include optimal quantile or expectile forecasts as a special case. The choice of the specification model reflects investigators' thoughts on the kind of deviation from this standard that seems plausible. The *break* model allows for structural change at a breakpoint $\theta_2 \in \mathbb{R}$, and the *linear* model admits linear dependence within the argument of the link function. The *periodic* model specifies the base level $\theta_0$, the amplitude $\theta_1$, and the period $\theta_2$ of any cyclic component. These models are continuous, except for the break model, which is discontinuous with respect to the breakpoint. However, in applications, there often exist natural choices for a fixed breakpoint (e.g., corresponding to a policy change or phase transition) and similarly for the period in the periodic model (e.g., implied by seasonal effects). Suitable adaptations apply when the state variable $s = (s_1, \ldots, s_k)'$ takes values in $\mathbb{R}^k$. For example, the linear model generalizes to $m(s, \theta) = \Phi(\theta_0 + \theta_1 s_1 + \ldots + \theta_k s_k)$. In any practical problem, the choice of the state variable and the parametric specification model will be guided and informed by substantive expertise, as we exemplify in economic and meteorological case studies. A plausible default, which we have found to be the most useful across a range of applications, uses the linear specification model with the probit link and the point forecast as the state variable.

## 3.2 | GMM estimator

In the following, we draw on standard GMM theory (Hansen, 1982; Newey & McFadden, 1986) to estimate specification models and test general hypotheses of optimality. We assume that $X_t$ is an optimal quantile forecast with $\mathcal{F}_t$-measurable state $S_t$ as in Equation (2). It follows that

$$\mathbb{E}[(\mathbb{1}(Y_t \leq X_t) - m(S_t, \theta_*))W_t] = 0 \quad \text{for all} \quad W_t \in \mathcal{F}_t.$$

In practice, we choose a $q$-variate instrument vector $w_t = (w_{t,1}, \ldots, w_{t,q})'$ that comprises information available to the forecast issuer at the time the forecast was made. We refer to

$$g_t(\theta) = (\mathbb{1}(y_t \leq x_t) - m(s_t, \theta))w_t$$

as the *moment function*. Alternatively, the forecast might be an optimal state-dependent expectile forecast, $X_t = e_{m(S_t,\theta_*)}(Y_t \mid \mathcal{F}_t)$, with associated moment function $g_t(\theta) = |1(y_t \leq x_t) - m(s_t, \theta)|(x_t - y_t)w_t$.

Given a sample path of instrument vectors $w_1, \ldots, w_T$, the empirical mean of the moment function is given by $g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta)$, and the GMM estimator $\hat{\theta}_T$ is obtained by minimizing the quadratic norm of $g_T(\theta)$, in that

$$\hat{\theta}_T = \arg\min_{\theta \in \Theta} g_T(\theta)' M_T g_T(\theta), \tag{3}$$

with a weighting matrix $M_T \in \mathbb{R}^{q \times q}$. Subject to customary assumptions, which include continuity of the specification model and sufficiently rich instruments to guarantee unique identifiability, the GMM estimator $\hat{\theta}_T$ in (3) is consistent. For formal statements, see Section S4.

Once consistency has been established, asymptotic laws in GMM theory can be applied. Throughout this study, we employ the standard two-step GMM procedure, where the second step relies on the inverse of the heteroskedasticity and autocorrelation consistent (HAC) noncentered covariance estimator $\Sigma_T$. Specifically, we follow Newey and West (1987) and use the linear Bartlett kernel and a sample-size-dependent bandwidth of $\lceil T^{1/5} \rceil$. Subject to regularity conditions that include continuous differentiability of the specification model and mixing conditions on the moment function, the two-step GMM estimator is asymptotically normal with

$$\sqrt{T}(\hat{\theta}_T - \theta_*) \to \mathcal{N}_p\left(0, (G\Sigma^{-1}G')^{-1}\right) \quad \text{as} \quad T \to \infty, \tag{4}$$

where $p$ is the dimension of the parameter vector, $\Sigma \in \mathbb{R}^{q \times q}$ is the covariance matrix of the moment function, and $G \in \mathbb{R}^{p \times q}$ is the expectation of its partial derivative with respect to $\theta$, at the true parameter value $\theta_*$. In the case of quantiles, we have $G = \mathbb{E}[m_{(\theta)}(S, \theta_*)W']$ and

$$\Sigma = \mathbb{E}[(\mathbb{1}(Y \leq X) - m(S, \theta_*))^2 WW'],$$

and in the case of expectiles, it holds that $G = \mathbb{E}[m_{(\theta)}(S, \theta_*) \, |Y - X| \, W']$ and

$$\Sigma = \mathbb{E}[(\mathbb{1}(Y \le X) - m(S, \theta_*))^2 (Y - X)^2 W W'].$$

For a formal treatment based on classical GMM theory, see Section S5.

While the GMM estimator is consistent for any selection of the instruments that guarantees identification, the associated tests of forecast optimality depend heavily on the instrument vector $W$. We tend to this issue now.

## 3.3 | Testing optimality with unknown directive

The well-known test of overidentifying restrictions (Hansen, 1982) can be used to test forecast optimality. Specifically, for state-dependent quantiles, our tests operate under the null hypothesis:

$$H_0 : \text{There exist } \theta_* \in \Theta \text{ and a } \sigma - \text{algebra } \mathcal{F} \text{ with } \sigma(W) \subseteq \mathcal{F} \text{ such that } X = q_{m(S,\theta_*)}(Y \mid \mathcal{F}), \tag{5}$$

and analogously for expectiles. Notably, the null hypothesis stipulates that the instrument vector $W$ is rationally included in the forecast and the specification model $m$ is correctly specified. As throughout we use specification models that nest optimal constant quantile or expectile forecasts, the approach allows for flexible, general hypotheses.

If the dimension $q$ of the instrument vector $W$ is greater than the dimension $p$ of the parameter vector $\theta$, and subject to the same regularity conditions as for the asymptotic distribution (4), the null hypothesis implies that

$$J_T(\hat{\theta}_T) \to \chi^2_{q-p} \quad \text{as} \quad T \to \infty,$$

where $J_T(\theta) = T \cdot g_T(\theta)' \, \Sigma_T^{-1} g_T(\theta)$ is called the $J$-statistic.

The choice of the instrument vector $W$ determines the information set for which we test. If a forecast is optimal with respect to $\mathcal{F}$, it also satisfies the moment conditions with respect to any smaller information set $\mathcal{G} \subseteq \mathcal{F}$. In particular, if a test with instrument vector $W$ rejects optimality, the point forecast is deemed suboptimal with respect to any information set $\mathcal{F}$ that contains $\sigma(W)$.

Suppose now that $X$ fails to be optimal under any parameter value $\theta_*$ and any $\sigma$-algebra $\mathcal{F}$. For a given instrument vector $W$, our optimality test has power against this global alternative, if the identification function remains correlated with the instruments. Subject to customary regularity conditions, the optimality test with instrument vector $W$ is consistent against

$$H_A : \mathbb{E}[g(\theta)] \neq \mathbf{0} \quad \text{for all} \quad \theta \in \Theta, \tag{6}$$

where the expectation is taken with respect to the joint distribution of $X$, $Y$, $S$, and $W$. Importantly, due to the finite dimension of the instrument vector, we test necessary (only) conditions for optimality. The inclusion of additional instruments increases asymptotic power but may degrade finite sample performance. The application of more flexible specification models allows to analyze more general classes of forecast optimality but decreases asymptotic power.

Optimality with respect to a specific information set can only be rejected if appropriate instruments are available to the evaluator. Furthermore, a misspecified or suboptimal forecast can still be optimal with respect to a smaller information set or a more flexible class of functionals. For these reasons, the choice of the specification model and of a sufficiently rich instrument vector is crucial. By its very nature, the current forecast value carries substantial information about the underlying quantile or expectile level, and we recommend strongly that it be included as a component of the instrument vector. The respective effects are illustrated in the simulation setting of Section 4.2 below.

The asymptotic results in Section 3.2 allow for the testing of general hypotheses about forecasting behavior in customary ways. For example, any restriction $R(\theta) = 0$ for the specification model $m(s, \theta)$, where $R : \Theta \mapsto \mathbb{R}^l$ is differentiable, can be tested for based on the Wald statistic (e.g., Greene, 2012).

# 4 | MONTE CARLO EXPERIMENTS

In this section, we demonstrate that our GMM estimator is reliable in finite-sample settings and induces well-calibrated and powerful tests, despite its flexibility. For ease of comparison to the related approach of Patton and

Timmermann (2007), which operates under the loss function paradigm, we adopt their simulation setting. Specifically, each sample path $y_1, \ldots, y_T$ is simulated from a stationary AR(1)–GARCH(1,1) model of the form

$$Y_t = \frac{1}{2} Y_{t-1} + \sigma_t \epsilon_t \quad \text{where} \quad \sigma_t^2 = \frac{1}{10} + \frac{4}{5}\sigma_{t-1}^2 + \frac{1}{10}\sigma_{t-1}^2 \epsilon_{t-1}^2, \tag{7}$$

with $\epsilon_t$ being standard Gaussian white noise. Let $\mathcal{I}_t$ be the filtration generated by the time series, $\mathcal{I}_t = \sigma(Y_t, Y_{t-1}, \ldots)$. We consider optimal forecasts based on distinct information sets in Section 4.1 and based on different specification models in Section 4.2. All results use a Monte Carlo sample of 2000 replicate time series, and all tests have nominal level 0.10. Code for replication is available at https://github.com/Schmidtpk/pf.

## 4.1 | State-independent forecasts under different information sets

We follow Patton and Timmermann (2007) and consider the point forecast

$$X_t = \frac{1}{2} Y_{t-1} - \frac{1}{4}\sigma_t. \tag{8}$$

Conditional on $\mathcal{F}_t = \mathcal{I}_{t-1}$, the point forecast $X_t$ is an optimal quantile forecast at the constant level $\tau = \Phi(-\frac{1}{4}) = 0.4012 \ldots$ Alternatively, the forecast can be interpreted as an optimal expectile forecast at the constant level $\tau = 0.3508 \ldots$ The two interpretations are equally valid, as the respective conditional distributions are all Gaussian. The same comment applies whenever the conditional distributions remain within a given location-scale family. See Yao and Tong (1996, proposition 1) and the discussion in Section S3.

We perform the overidentifying restrictions test of forecast optimality from Section 3.3 and compare tests based on state-dependent quantiles with distinct specification models in the state variable $X_t$, all of which nest the case of a fixed quantile, to the flexible spline test with state variable $Y_t$ as specified in eq. (16) of Patton and Timmermann (2007). Specifically, we consider the linear specification model and the constant specification model, which yields a test of the type proposed by Elliott et al. (2005). The quantile models both use $W_t = (1, X_t, Y_{t-1})'$, and the spline approach uses the instrument vector $W_t = (1, X_t, Y_{t-1}, X_t^2, Y_{t-1}^2, X_t^3, Y_{t-1}^3)'$, which generates the same $\sigma$-algebra. Figure 3 demonstrates that the quantile-based tests are much better calibrated than the flexible spline test, which is oversized throughout.

A possible reason is that, using a single node (at zero) only, the flexible spline test reduces to the expectile test with a linear-logistic specification model and an inadmissible (i.e., not $\mathcal{I}_{t-1}$-measurable) state variable $Y_t$, which seems problematic from both theoretical and substantive points of view.

It is interesting to observe that in the original loss function formulation the spline approach seems entirely innocuous, with measurability issues not being apparent at all. Considering the spline with admissible state variables, such as $Y_{t-1}$ or $X_t$, this critique does not apply, but the general identifiability problem of the loss function approach as described in Section 2 persists. Indeed, in a range of simulation settings both with and without state dependence (in experiments not shown here), the spline-based estimator appears to be unidentified and inconsistent. In contrast, the approach based on state-dependent quantiles and expectiles provides insightful point estimates, reliable confidence intervals, and nearly nominally sized tests.

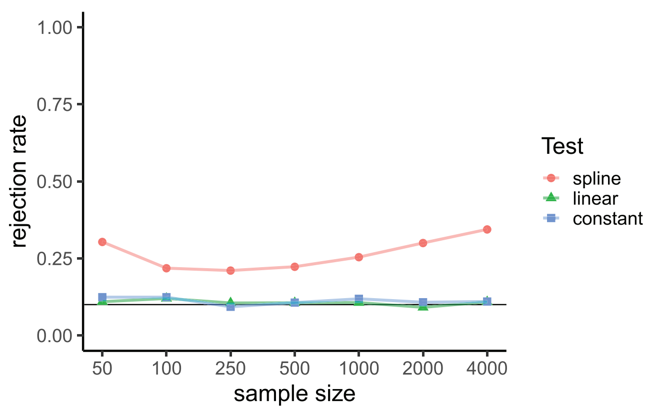


**FIGURE 3** Size of optimality tests for the one-step-ahead quantile forecast. The horizontal line is at the nominal level of 0.10 [Colour figure can be viewed at wileyonlinelibrary.com]

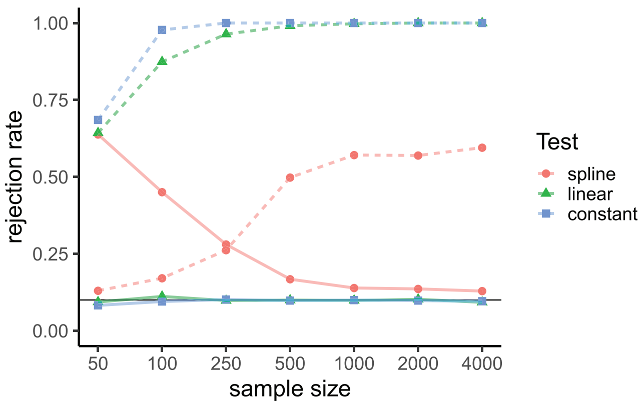


**FIGURE 4** Size and size-adjusted power of expectile-based optimality tests for the two-step-ahead mean forecast. The horizontal line is at the nominal level of 0.10. The solid lines represent size for tests with properly lagged instruments. The dashed lines represent size-adjusted power for tests with nonlagged instruments [Colour figure can be viewed at wileyonlinelibrary.com]

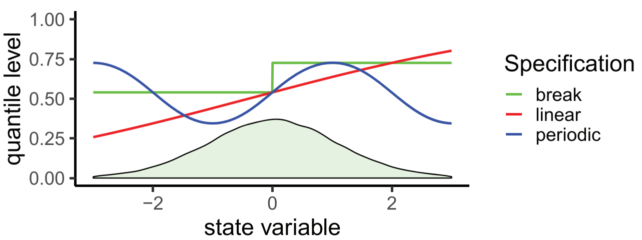


**FIGURE 5** Quantile level as a function of the state variable under the specification models in Section 4.2. The density of the state variable $Y_{t-1}$ is also shown [Colour figure can be viewed at wileyonlinelibrary.com]

In the subsequent analysis, we study power against information rigidities (Coibion & Gorodnichenko, 2015) and construct a two-step-ahead forecast with lagged information set $\mathcal{F}_t = \mathcal{I}_{t-2}$. The respective conditional predictive distributions have mean and median $\mathbb{E}[Y_t \mid \mathcal{I}_{t-2}] = q_{1/2}(Y_t \mid \mathcal{I}_{t-2}) = \frac{1}{4}Y_{t-2}$, as we show in Section S6. However, $\frac{1}{4}Y_{t-2}$ fails to be the optimal mean or median forecast with respect to $\mathcal{I}_{t-1}$. A well-performing test rejects optimality based on instruments in $\mathcal{I}_{t-1}$ and obtains the nominal value with information in $\mathcal{I}_{t-2}$. We compare expectile-based optimality tests to the flexible spline test, and in the former case, we use the same instrument vectors as before, whereas in the latter case, we substitute $Y_{t-1}$ by the valid instrument $Y_{t-2}$. As shown in Figure 4, the expectile-based tests are better calibrated and more powerful than the spline-based test, which is strongly oversized for small sample sizes and unable to consistently detect the information rigidity. The loss of power for the state-dependent test relative to the less flexible constant test is modest. Additional results and discussion on estimation and testing in Section S7 cover quantile-based tests and generally support these findings.

## 4.2 | State-dependent forecasts under different specification models

Next, we investigate whether our optimality test can discriminate among the proposed specification models. To this end, we generate optimal state-dependent forecasts for the data-generating process (7) with the most recent outcome of the time series, $Y_{t-1}$, as state variable. Specifically, we let

$$X_t = \frac{1}{2}Y_{t-1} + q_{m(Y_{t-1})}(\mathcal{N}(0,1))\sigma_t,$$

where $m(s) = \Phi(\frac{1}{10} + \frac{s}{4})$ under the linear specification model, $m(s) = \Phi(\frac{1}{10} + \frac{1}{2} \cdot \mathbb{1}(s \geq 0))$ under the break model, and $m(s) = \Phi\left(\frac{1}{10} + \frac{1}{2}\sin\left(\frac{\pi}{2}s\right)\right)$ under the periodic model, as illustrated in Figure 5. We consider specification models with two parameters only; to achieve this, we fix the breakpoint in the break model and the period in the periodic model at their respective true values.

We then apply overidentifying restrictions tests of forecast optimality with instrument vector $W_t = (1, Y_{t-1}, X_t)'$. Table 2 shows results for forecasts and tests based on quantiles. Even for small sample sizes, our tests are reasonably calibrated and quite powerful with rejection rates up to 70% for more distinct specifications. For larger sample sizes, the optimality tests are almost perfectly calibrated with high power.

As noted, the test depends crucially on the choice of the instrument vector. The final block of columns in Table 2 considers the same setting as before, except for the instruments used. Specifically, we drop the forecast value and now use a constant and the first two lagged outcomes of the time series, so that $W_t = (1, Y_{t-1}, Y_{t-2})'$. While the tests continue

**TABLE 2** Rejection rates of optimality tests based on quantile specification models with state variable $Y_{t-1}$

| Sample size | True model | Hypothesized model $W_t = (1, Y_{t-1}, X_t)'$ Linear | Break | Periodic | $W_t = (1, Y_{t-1}, Y_{t-2})'$ Linear | Break | Periodic |
|---|---|---|---|---|---|---|---|
| $T = 100$ | Linear | 0.10 | 0.23 | 0.36 | 0.10 | 0.09 | 0.09 |
| | Break | 0.28 | 0.11 | 0.08 | 0.10 | 0.09 | 0.09 |
| | Probit | 0.65 | 0.30 | 0.07 | 0.09 | 0.09 | 0.09 |
| $T = 250$ | Linear | 0.11 | 0.42 | 0.87 | 0.11 | 0.11 | 0.11 |
| | Break | 0.52 | 0.10 | 0.15 | 0.10 | 0.11 | 0.10 |
| | Probit | 0.96 | 0.58 | 0.09 | 0.09 | 0.09 | 0.09 |
| $T = 1000$ | Linear | 0.11 | 0.86 | 1.00 | 0.11 | 0.12 | 0.13 |
| | Break | 0.96 | 0.10 | 0.44 | 0.10 | 0.10 | 0.10 |
| | Probit | 1.00 | 0.99 | 0.10 | 0.10 | 0.10 | 0.10 |

*Note:* The nominal level is 0.10, and the test uses the instrument vector $W_t = (1, Y_{t-1}, X_t)'$ or $W_t = (1, Y_{t-1}, Y_{t-2})'$, respectively.

to be well calibrated, power against the alternatives vanishes. In contrast to $Y_{t-2}$, the forecast $X_t$ is a major contributor of further information, and the test rejects optimality with respect to the larger information set. We generally recommend that the forecast value, which by its very nature carries information about the underlying forecast directive, be included in the instrument vector. Section S8 considers the case where the state variable in the linear quantile specification model is the point forecast $X_t$, rather than the most recent outcome, and Section S9 provides further results in the case where expectile specification models are hypothesized in lieu of quantile models, with similar findings.

## 5 | EMPIRICAL EXAMPLES

In this section, we provide economic and meteorological case studies. In both cases, optimality of the point forecast with respect to a constant quantile or expectile level is rejected, but optimality with respect to more flexible state-dependent directives cannot be rejected. We illustrate that the estimated specification models yield accessible and relevant information for forecasters and forecast users. Unless noted otherwise, we apply the linear specification model with the probit link function and the current forecast value as state variable, and the instrument vector includes a constant, the forecast value at hand, and the most recent outcome available when the forecast was issued.

### 5.1 | GDP growth forecasts as state-dependent quantiles

We revisit the well-studied Greenbook forecasts of the US Federal Reserve for GDP growth. As multiple forecasts are issued within a quarter for the GDP growth in the next quarter, we consider two different one-quarter-ahead forecasts. Specifically, for each GDP observation, we consider the forecast issued closest to the midpoint (main forecast) and closest to the end (late forecast) of the previous quarter. As realized values, we take the quarterly real GDP growth rate in the United States over the period 1969 to 2012 for $T = 176$ observations, as reported in the initial data release and illustrated in Figure 1.[1] In a pioneering effort, Patton and Timmermann (2007) modeled the Federal Reserve's loss function as a quadratic spline with three nodes whose shape is allowed to change with the realized growth rate.

Here, we interpret the forecasts as quantiles of the Federal Reserve's (implicit) predictive distributions. As instrumental variables for the GMM estimator, we employ a constant, the most recent outcome available at the time of the forecast, and the one-step-ahead forecast at hand, that is, $w_t = (1, y_{t-2}, x_t)'$. For the late forecast issued at the end of the previous quarter, there is no evidence against the hypothesis of an optimal quantile forecast at a constant level. The estimated quantile level is 0.59 with standard error 0.04, and the $p$-value in the associated test of overidentifying restrictions is 0.30. Next, we implement the Wald test introduced in Section 3.3. The hypothesis of an optimal median forecast ($\theta_0 = \frac{1}{2}$) is rejected with a $p$-value of 0.02, so forecast users will keep in mind that an optimal median forecast would be consistently smaller than the value issued.

For the main forecast, the hypothesis of an optimal quantile forecast at a constant level is untenable, as the $p$-value in the test of overidentifying restrictions drops to 0.07. To investigate whether the main forecast is optimal if we allow the reported quantiles to change with the predicted GDP growth rate, $x_t$, we apply the linear specification model $m(x_t, \theta) = \Phi(\theta_0 + \theta_1 x_t)$. Compared with the spline loss function of Patton and Timmermann (2007), we use two only instead of six

[1]The results in this section are robust to using the second revision or the most recent vintage.

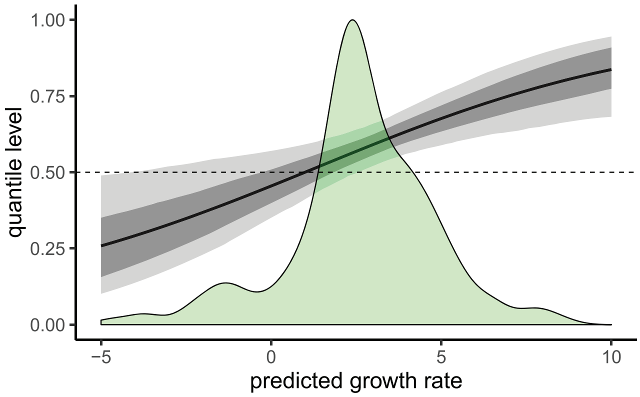


**FIGURE 6** Estimated quantile specification model for the Federal Reserve's forecasts of gross domestic product (GDP) growth plotted against the predicted growth rate (in percent), with pointwise confidence intervals at level 0.60 and 0.90. A density estimate of the point forecast is also shown [Colour figure can be viewed at wileyonlinelibrary.com]

parameters, and we apply state variables and instruments that are at least implicitly available at the time when the forecast is issued. The test of overidentifying restrictions yields a $p$-value of 0.49.[2]

The GMM estimate for $\theta = (\theta_0, \theta_1)'$ is $\hat{\theta}_T = (-0.10, 0.11)'$, and we interpret the forecasts as $m(x_t, \hat{\theta}_T)$-quantiles that depend on the predicted growth rate $x_t$. The covariance estimate implied by (4) is given by

$$\frac{1}{T}\left(\hat{G}_T' \hat{\Sigma}_T^{-1} \hat{G}_T\right)^{-1} = \begin{pmatrix} 0.028 & -0.006 \\ -0.006 & 0.002 \end{pmatrix},$$

where $\hat{G}_T$ is a HAC estimator for $G$. As $\Phi$ is strictly monotone, we can compute pointwise confidence intervals for $\hat{\theta}_0 + \hat{\theta}_1 x_t$ and transform into confidence intervals for $m(x_t, \hat{\theta}) = \Phi(\hat{\theta}_0 + \hat{\theta}_1 x_t)$, as illustrated in Figure 6. The Federal Reserve reports higher quantile levels during times of strongly positive expected growth. The model also stipulates lower quantile levels in times of negative expected growth, but the confidence bands do not exclude the median forecast except for extreme cases. A Wald test of the hypothesis that the forecaster's behavior does not change with respect to the state variable ($\theta_1 = 0$) is rejected with a $p$-value of 0.01. Thus, for the forecast to be rational, the issued quantile needs not only be asymmetric but also be adaptive to the state variable. Similarly parsimonious and simultaneously informative summaries of inference results would not be attainable with extant approaches.

The Greenbook projection is a judgmental forecast based on quantitative data, qualitative information, economic models, and a range of forecasting techniques (Edge et al., 2010), and the estimated specification model admits potential explanations in various directions. As noted by Patton and Timmermann (2007), the Federal Reserve staff might exhibit preferences that depend on the outlook itself. Alternatively, the estimated relationship is consistent with overreacting to information (or overfitting). Finally, forecasters might underestimate the impact or severity of measures enacted by the Federal Reserve or the government. If enacted policies push growth rates more heavily toward the long-time trend than anticipated, the observed pattern emerges, where severe expected growth rates are not in line with equally severe realizing growth rates. While further studies are beyond the scope of our paper, we believe that our findings can help staff at the Federal Reserve Board understand and improve their projections. Conversely, forecast users will appreciate that projected values in the tails should not be understood as central tendencies and may use the point estimates in Figure 6 for a state-dependent interpretation of the main forecast.

We summarize that it is key to consider the available information in tests of forecast optimality. While the late Greenbook forecasts issued toward the end of the quarter are optimal in the classical sense, the driving factor of overly optimistic forecasting in the main forecast earlier in the quarter is anticipated growth, whereas the behavior in anticipation of recessions is inconclusive.

## 5.2 | Precipitation forecasts as state-dependent expectiles

Numerical weather prediction has seen tremendous advances over the past decades (Bauer et al., 2015). Here, we consider 24-h-ahead forecasts of daily precipitation accumulation over London, UK, from the high-resolution run operated by the ECMWF (www.ecmwf.int). While the forecasts are generated by numerical models that are run on supercomputers

[2]If we apply the same specification model, but use the lagged outcome $y_{t-2}$ instead of the forecast $x_t$ as the state variable, forecast optimality is rejected with $p$-value 0.028. Similarly, if we use the lagged outcome $y_{t-1}$ as the state variable and change the instrument vector to $(1, y_{t-1}, x_t)'$, forecast optimality is rejected with $p$-value 0.046.

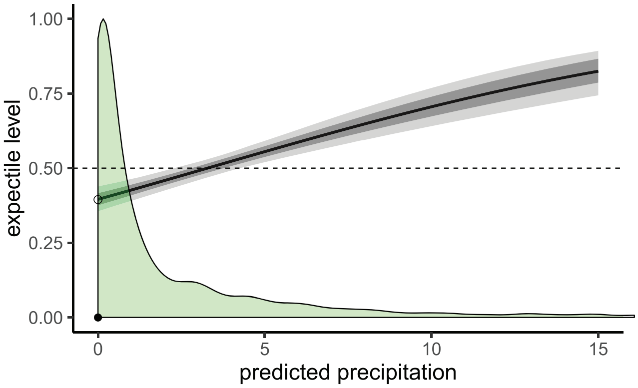


**FIGURE 7** Estimated expectile specification model for European Centre for Medium-Range Weather Forecasts (ECMWF) forecasts of daily rainfall at London plotted against the predicted precipitation (in mm), with pointwise confidence intervals at level 0.60 and 0.90. The dots illustrate the discontinuity of the expectile level at a point forecast of $x_t = 0$, as specified in Equation (9). A density estimate of the point forecast conditional on it being strictly positive is also shown [Colour figure can be viewed at wileyonlinelibrary.com]

(ECMWF Directorate, 2012), many facets of the operational implementation are subject to tuning by human experts, often based on decade-long experience. In our analysis, we use ECMWF forecasts and observations from the ERA-Interim system (Dee et al., 2011) in the period July 1, 2011, to June 30, 2017, for $T = 2192$ observations, as partly illustrated in Figure 2. Despite good agreement between forecasts and observations, the hypothesis of an optimal constant expectile forecast is rejected with a $p$-value $< 10^{-8}$. Instead, we interpret the forecasts as state-dependent expectiles of the underlying predictive distribution. We employ expectiles rather than quantiles to avoid artifacts due to the mixed discrete-continuous nature of precipitation accumulation, which is a nonnegative variable with a point mass at zero. Any point forecast $x_t = 0$ can be interpreted as essential infimum or expectile at level $\tau = 0$. To investigate whether the reported expectile varies with predicted precipitation accumulations $x_t > 0$, we apply a slightly modified linear specification model,

$$m(x_t, \theta) = \Phi(\theta_0 + \theta_1 x_t)\,\mathbb{1}(x_t > 0). \tag{9}$$

The GMM estimate with instrument vector $w_t = (1, y_{t-2}, x_t)'$ for $\theta = (\theta_0, \theta_1)'$ is $\hat{\theta}_T = (-0.27, 0.08)'$. These results are robust to using the instrument vector $w_t = (1, y_{t-1}, x_t)'$. However, as $y_{t-1}$ represents the daily precipitation accumulation, it has not yet been observed when the numerical model that generates $x_t$ is initialized. The resulting specification model is illustrated in Figure 7, and the test of overidentifying restrictions does not reject forecast optimality, with the $p$-value being at 0.74.[3] It is interesting to observe that the numerical model generates considerably higher expectile levels in anticipation of severe, extreme rain.

These results invite speculation about underlying reasons, particularly in view of the inevitable tuning of the numerical model by human experts. One possible explanation is via the forecasters's dilemma (Lerch et al., 2017), which refers to the fact that the public is most concerned about forecast performance in cases of extreme events. Hence, human experts may feel implicit incentives to overstate in anticipation of severe storms. Alternatively, geoscientists occasionally express a belief that the model climate—that is, the marginal distribution of the forecast—ought to equal the observed climate—that is, the marginal distribution of the outcomes. As is easily appreciated, point forecasts at a constant quantile or expectile level are incompatible with the equality of actual and model climate. However, the desired approximate equality is attainable under relationships of the type seen in Figure 7, where forecasts are issued at increasingly higher, state-dependent levels. Clearly, there is a need and scope for further investigation, including but not limited to analyses at locations worldwide, and the reasons for observed behaviors will very likely be multifaceted. In particular, they are bound to include location-dependent interactions between model settings—referred to as parameterizations by meteorologists—and local and regional specifics in terrain and atmospheric conditions.

Both forecast developers and forecast users can benefit from these findings. For forecast developers and their organizations, they provide feedback that can be used to inform model updates. Likewise, local forecast users will appreciate that point forecasts above 5 mm dominate median forecasts, and broadcast meteorologists in London will be wise to avoid further exaggeration when communicating extreme forecasts to their audiences.

[3]The use of the state variable $y_{t-2}$ yields a $p$-value $< 10^{-8}$.

## 6 | DISCUSSION

For point forecasts with unknown directive, we posit that it is preferable to estimate and test the functional quoted by the forecaster, rather than the loss function, for reasons of identifiability, interpretability, and ease and efficiency of inference. While our approach deviates from recent literature, the classical Mincer and Zarnowitz (1969) test of forecast optimality can be interpreted in our setting. Specifically, under the regression model $y_t = \beta_0 + \beta_1 x_t + u_t$, tests of the coefficients being equal to zero and one, respectively, can be interpreted as assuming an optimal forecast of the functional form $\beta_0+\beta_1\mathbb{E}[Y|X]$ with identification function $V(x, y) = \beta_0 + \beta_1 x - y$, and applying the GMM estimator with instruments $w = (1, x)'$, with ensuing tests.

Lieli and Stinchcombe (2013) show that some classes of loss functions are theoretically identifiable if the eligible set of forecasts is restricted. Under sufficient variation in the arising conditional distributions, loss functions that are consistent for an optimal mean forecast are identifiable. However, the class of loss functions consistent for a specific optimal quantile forecast cannot be identified (Lieli et al., 2019). State-dependent quantiles and expectiles are identifiable irrespectively of these types of assumptions.

Under misspecification or nonnested information sets, forecast comparison is not robust to the choice of loss function even within the class of loss functions defining the same optimal forecast (Patton, 2020). It is particularly noteworthy that state-dependent functionals allow for the treatment of supposedly misspecified forecasts in a principled manner. While dominance across all loss functions can be assessed, this feature may be uncommon in practice (Ehm et al., 2016). If forecast comparison is inconsistent across loss functions, the methods provided here might be useful to assess miscalibration and the validity of specific information sets. In particular, identifying moment conditions can be used to test if a forecast is optimal with respect to an information set that encompasses a second forecast.

We have introduced quantile and expectile specification models for the description of forecasting behavior in terms of a state variable, thereby relating to extant work on time-varying quantiles and expectiles (De Rossi & Harvey, 2009). Under the assumption of optimal forecasts, the model parameters can be consistently estimated, and the asymptotic distributions of the GMM estimator and the respective test statistics can be used to construct flexible tests of forecast optimality and specific model properties. Furthermore, support conditions and mixed discrete-continuous distributions can be handled efficiently and with rigor, as illustrated in the precipitation example. An accompanying software package `PointFore` in R (R Core Team, 2018) is available on CRAN.

For valid estimation and testing, the state variables need to be in the information set of the forecaster. A valid and universally available state variable is the forecast value at hand. The GMM estimator depends on an appropriate choice of a sufficiently rich instrument vector, and we recommend that the forecast value at hand be included as an instrumental variable. Importantly, the null hypothesis in the overidentifying restrictions test of forecast optimality reflects the selection of the instrument vector, and any specific choice yields necessary (but not sufficient) conditions for forecast optimality. Judicious choices are critical, as we have demonstrated in real data and simulation examples.

We define optimality as a property of the observable forecast or more specifically as a parametric relation between the forecast and the ideal predictive distribution. Similar settings can be found in Giacomini and White (2006) for forecast comparison via loss differences and in Giacomini and Komunjer (2005) for quantile encompassing tests. This strong hypothesis of optimality enables the use of standard tests and yields context independent asymptotic distributions. In a more elaborate approach, this assumption is substituted for by specific assumptions on the forecast generation, for example, via a correctly specified time series model and efficient estimation in an increasing training window, thereby accounting for estimation uncertainty (Elliott et al., 2005; Guler et al., 2017). Reassuringly, in the case of a constant specification model, our asymptotic results are identical to extant results that account for estimation uncertainty assuming a super-consistent estimator (Elliott et al., 2005, proposition 4). However, in unstable environments, our optimality assumption may be too strong and a more elaborate description that accounts for estimation uncertainty may be warranted (West, 1996).

In empirical examples from economics and meteorology, we have applied the linear specification model with the forecast at hand as state variable, and we have found common ground, in that the forecasts are indicative of overstating in anticipation of extreme events. While application-specific reasons for this remain to be explored in further detail, a potential interpretation is via the forecasters's dilemma (Lerch et al., 2017), which refers to the fact that the public's attention focuses on the forecast performance in cases of extreme events. As a result of this ubiquitous practice, individual and institutional forecasters may have implicit incentives to exaggerate in anticipation of tail events. A related phenomenon is the hard–easy effect described in the psychological literature, in that human subjects tend to be overconfident in answering hard questions, while being underconfident in responses to easy questions (Kynn, 2008; Lichtenstein et al., 1982).

Further investigation is called for in order to assess whether overstatement in anticipation of extreme events is characteristic when predictions are generated by individual or institutional forecasters, or when numerical models are designed, tuned, and informed by human expertise. The commonalities between our two unlike examples from economics and meteorology suggest that the linear specification model with the forecast value as state variable might be a useful default choice in studying this type of question.

Generally, forecasters and their institutions may use our methods to evaluate past projections and check whether the findings are consistent with internal objectives and react by adapting forecasting procedures, if necessary. For example, rejections of forecast optimality based on information generated by an instrument indicate that said information has not been consistently incorporated. A state-dependent model with said instrument as state variable can then be used to understand the nature of the deviation. Likewise, forecast users might be interested in using our approach, for guidance in the use and communication of the forecasts.

## ACKNOWLEDGMENTS

The authors thank the European Centre for Medium-Range Weather Forecasts (ECMWF) in Reading, UK, for providing meteorological data and are grateful to Stephan Hemri for assistance in their handling. The work of Patrick Schmidt, Tilmann Gneiting, and Stephan Hemri was partially funded by the Klaus Tschira Foundation and by the European Union Seventh Framework Programme under Grant 290976. Tilmann Gneiting also is grateful for travel support and encouragement through the ECMWF Fellowship Programme. Matthias Katzfuss was partially supported by US National Science Foundation (NSF) Grant DMS–1521676 and NSF CAREER Grant DMS–1654083. Further, the authors are grateful to the co-editor, three anonymous referees, Werner Ehm, Tobias Fissler, Alexander Glas, Peter Knippertz, Fabian Krüger, Barbara Rossi, Michael Scheuerer, and Peter Vogel for a wealth of constructive and insightful comments.

## OPEN RESEARCH BADGES

This article has been awarded Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. Data is available at https://urldefense.com/v3/__http://qed.econ.queensu.ca/jae/datasets/schmidt001/_;!!N11eV2iwtfs!6j4_zQDGzI8sbkUY6gwZsPRSGkClzhGKs1LQDigJ-b40vBc_o23m3njqu-KcTRGz$.

## REFERENCES


Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*, 47–55.

Bellini, F., Klar, B., Müller, A., & Gianin, E. R. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, *54*, 41–48.

Billingsley, P. (1995). *Probability and measure* (3rd ed.). John Wiley & Sons.

Capistrán, C. (2008). Bias in Federal Reserve inflation forecasts: Is the Federal Reserve irrational or just cautious? *Journal of Monetary Economics*, *55*, 1415–1427.

Christodoulakis, G. A., & Mamatzakis, E. C. (2008). An assessment of the EU growth forecasts under asymmetric preferences. *Journal of Forecasting*, *27*, 483–492.

Coibion, O., & Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, *105*, 2644–2678.

De Rossi, G., & Harvey, A. (2009). Quantiles, expectiles and splines. *Journal of Econometrics*, *152*, 179–185.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., & Thépaut, J. N. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*, 553–597.

Diebold, F. X., & Lopez, J. A. (1996). Forecast evaluation and combination. In Maddala, G. S., & Rao, C. R. (Eds.), *Statistical methods in finance*, Handbook of Statistics, Vol. *14*: Elsevier, pp. 241–268.

ECMWF Directorate (2012). Describing ECMWF's forecasts and forecasting system. *ECMWF Newsletter*, *133*, 11–13.

Edge, R. M., Kiley, M. T., & Laporte, J.-P. (2010). A comparison of forecast performance between Federal Reserve staff forecasts, simple reduced-form models, and a DSGE model. *Journal of Applied Econometrics*, *25*, 720–754.

Ehm, W., Gneiting, T., Jordan, A., & Krüger, F. (2016). Of quantiles and expectiles: Consistent scoring functions, Choquet representations and forecast rankings (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *78*, 505–562.

Elliott, G., Komunjer, I., & Timmermann, A. (2005). Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies*, *72*, 1107–1125.
Elliott, G., Komunjer, I., & Timmermann, A. (2008). Biases in macroeconomic forecasts: Irrationality or asymmetric loss? *Journal of the European Economic Association*, *6*, 122–157.
Elliott, G., & Timmermann, A. (2008). Economic forecasting. *Journal of Economic Literature*, *46*, 3–56.
Engelberg, J., Manski, C. F., & Williams, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business & Economic Statistics*, *27*, 30–41.
Fissler, T., & Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *Annals of Statistics*, *44*, 1680–1707.
Fritsche, U., Pierdzioch, C., Rülke, J.-C., & Stadtmann, G. (2015). Forecasting the Brazilian real and the Mexican peso: Asymmetric loss, forecast rationality, and forecaster herding. *International Journal of Forecasting*, *31*, 130–139.
Giacomini, R., & Komunjer, I. (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business & Economic Statistics*, *23*, 416–431.
Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*, 1545–1578.
Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, *106*, 746–762.
Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, *1*, 125–151.
Greene, W. H. (2012). *Econometric analysis* (7th ed.). Prentice Hall.
Guler, K., Ng, P. T., & Xiao, Z. (2017). Mincer–Zarnowitz quantile and expectile regressions for forecast evaluations under asymmetric loss functions. *Journal of Forecasting*, *36*, 651–679.
Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*, 1029–1054.
Heinrich, C. (2014). The mode functional is not elicitable. *Biometrika*, *101*, 245–251.
Holzmann, H., & Eulert, M. (2014). The role of the information set for forecasting—With applications to risk management. *Annals of Applied Statistics*, *8*, 595–621.
Horowitz, J. L., & Manski, C. F. (2006). Identification and estimation of statistical functionals using incomplete data. *Journal of Econometrics*, *132*, 445–459.
Huber, P. J., & Ronchetti, E. M. (2009). *Robust statistics* (2nd ed.). John Wiley & Sons.
Krol, R. (2013). Evaluating state revenue forecasting under a flexible loss function. *International Journal of Forecasting*, *29*, 282–289.
Kynn, M. (2008). The ‘heuristics and biases’ bias in expert elicitation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *171*, 239–264.
Lerch, S., Thorarinsdottir, T. L., Ravazzolo, F., & Gneiting, T. (2017). Forecaster's dilemma: Extreme events and forecast evaluation. *Statistical Science*, *32*, 106–127.
Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In Kahnemann, D., Slovic, P., & Tversky, A. (Eds.), *Judgement under uncertainty: Heuristics and biases*: Cambridge University Press, pp. 306–334.
Lieli, R. P., & Stinchcombe, M. B. (2013). On the recoverability of forecasters' preferences. *Econometric Theory*, *29*, 517–544.
Lieli, R. P., Stinchcombe, M., & Grolmusz, V. (2019). Unrestricted and controlled identification of loss functions: Possibility and impossibility results. *International Journal of Forecasting*, *35*, 878–890.
Manski, C. F. (2016). Interpreting point predictions: Some logical issues. *Foundations and Trends in Accounting*, *10*, 238–261.
Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In Mincer, J. A. (Ed.), *Economic forecasts and expectations: Analysis of forecasting behavior and performance*: National Bureau of Economic Research, pp. 3–46.
Newey, W. K., & McFadden, D. (1986). Large sample estimation and hypothesis testing. In Engle, R. F., & McFadden, D. (Eds.), *Handbook of econometrics*, Vol. *4*: Elsevier, pp. 2111–2245.
Newey, W. K., & Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, *55*, 819–847.
Newey, W. K., & West, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*, 703–708.
Patton, A. J. (2020). Comparing possibly misspecified forecasts. *Journal of Business & Economic Statistics*, *38*, 796–809.
Patton, A. J., & Timmermann, A. (2007). Testing forecast optimality under unknown loss. *Journal of the American Statistical Association*, *102*, 1172–1184.
Patton, A. J., & Timmermann, A. (2010). Generalized forecast errors, a change of measure, and forecast optimality conditions. In Bollerslev, T., Russell, J., & Watson, M. (Eds.), *Volatility and time series econometrics: Essays in honor of Robert Engle*: Oxford University Press, pp. 194–212.
Pierdzioch, C., Rülke, J.-C., & Stadtmann, G. (2013). Oil price forecasting under asymmetric loss. *Applied Economics*, *45*, 2371–2379.
R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, Available at https://www.R-project.org/
Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, *66*, 783–801.
Steinwart, I., Pasin, C., Williamson, R., & Zhang, S. (2014). Elicitation and identification of properties. In *Conference on learning theory 2014* (Balcan, M. F., & Szepesvári, C., Eds.), Proceedings of Machine Learning Research, *35*, pp. 482–526.
Wang, Y., & Lee, T.-H. (2014). Asymmetric loss in the Greenbook and the Survey of Professional Forecasters. *International Journal of Forecasting*, *30*, 235–245.
West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, *64*, 1067–1084.
Yao, X., & Tong, H. (1996). Asymmetric least squares estimation: A nonparametric approach. *Nonparametric Statistics*, *6*, 273–292.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Schmidt, P., Katzfuss M., & Gneiting T. (2021). Interpretation of point forecasts with unknown directive. *Journal of Applied Econometrics*. 1–16. https://doi.org/10.1002/jae.2833