Supplementary Material: Interpretation of Point Forecasts with Unknown Directive

Patrick Schmidt*
University of Zurich, Zurich, Switzerland

Matthias Katzfuss
Texas A&M University, College Station, USA

Tilmann Gneiting
Heidelberg Institute for Theoretical Studies, Heidelberg, Germany
Karlsruhe Institute of Technology, Karlsruhe, Germany

December 6, 2020

* *Address for correspondence:* Patrick Schmidt, University of Zurich, Rämistrasse 71, 8006 Zürich, Switzerland. E-mail: patrickwolfgang.schmidt@uzh.ch. 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 agreement no. 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, Fabian Krüger, Barbara Rossi and Peter Vogel for a wealth of constructive and insightful comments.

S1 Identifying moment conditions

Consider the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where the elements of the sample space Ω are tuples that comprise the point forecast X, the realization Y, and a covariate vector Z. We assume that the information set $\mathcal{F}$ is a sub-σ-algebra of $\mathcal{A}$. If no measure is explicitly mentioned, statements like almost surely refer to $\mathbb{P}$. For random variables R_1 and R_2, we simply write $R_1 = R_2$ instead of $R_1 = R_2$ almost surely. In particular, statements like $X = \alpha(Y \mid \mathcal{F})$ denote $\mathbb{P}$-almost sure properties. As defined in Section 2 of the main paper standard measurability and integrability conditions are denoted $R \in \mathcal{F}_Q$ and $R \in \mathcal{F}$.

Before proceeding to our main results on identifying moment conditions, we state an elementary measure-theoretic equivalence.

Lemma 1 *For every integrable random variable U,*

$$\mathbb{E}(U \mid \mathcal{F}) = 0 \iff \mathbb{E}[UW] = 0 \ \textit{for all} \ W \in \mathcal{F}_U.$$

Proof: The implication from left to right is immediate from Theorem 34.3 in Billingsley (1995). For the reverse implication let W be the indicator function of any event A in the information set $\mathcal{F}$, to yield $\int_A \mathbb{E}(U \mid \mathcal{F}) \, \mathrm{d}\mathbb{P} = 0$ for all $A \in \mathcal{F}$, which implies $\mathbb{E}(U \mid \mathcal{F}) = 0$ by a standard argument. □

We now consider the τ-quantile functional q_τ and the τ-expectile functional e_τ, which includes the special case $\tau = \frac{1}{2}$ of the mean-functional. The following assumptions ensure that the functional is single-valued and well defined.

- A_Q The conditional distribution $\mathcal{L}(Y \mid \mathcal{F})$ is absolutely continuous with a strictly increasing cumulative distribution function almost surely.

- A_E The conditional distribution $\mathcal{L}(Y \mid \mathcal{F})$ has finite mean and positive variance almost surely.

Lemma 2 (quantiles) *Under condition A_Q the function $V_\tau(x, y) = \mathbb{1}(x \geq y) - \tau$ identifies the optimal τ-quantile forecast, i.e.,*

$$X = q_\tau(Y \mid \mathcal{F}) \iff \mathbb{E}[(\mathbb{1}(X \geq Y) - \tau)W] = 0 \ \textit{for all} \ W \in \mathcal{F}.$$

Proof: For every $\omega \in \Omega$ the definition of the τ-quantile implies that $x = q_\tau(Y \mid \mathcal{F})(\omega) \iff \mathbb{E}[V_\tau(x, Z)] = 0$, where $Z \sim \mathcal{L}(Y \mid \mathcal{F})(\omega) \in \mathcal{P}$. In terms of the $\mathcal{F}$-measurable random variable X this equality can be stated as $X = q_\tau(Y \mid \mathcal{F}) \iff \mathbb{E}[V_\tau(X, Y) \mid \mathcal{F}] = 0$. The stated equivalence is now immediate from Lemma 1. □

Lemma 3 (expectiles) *Under condition A_E the function $V_\tau(x, y) = |\mathbb{1}(x \geq y) - \tau|(x - y)$ identifies the optimal τ-expectile forecast, i.e.,*

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

The proof is essentially the same as in the case of quantiles. Very similar results are stated in Section 2 of Elliott et al. (2005) under the further assumption that X is a linear function in W.

The next and final result in this section is a variant of findings in Steinwart et al. (2014), and we follow the terminology used in their paper. In particular, topological statements on the space of probability distributions with bounded Lebesgue measures are with respect to the metric induced by the L_1-norm. The conditions on the functional α are met if it is defined via a continuous, nontrivial loss function, for continuity follows from the Maximum Theorem (e.g., Ok 2007, p. 229), and functionals defined via loss functions have convex level sets (Osband 1985, Gneiting 2011).

Lemma 4 *Let $\mathcal{P}$ be a convex set of probability measures with bounded Lebesgue densities such that $\mathcal{L}(Y \mid \mathcal{F}) \in \mathcal{P}$ almost surely, and suppose that the functional $\alpha : \mathcal{P} \mapsto \mathbb{R}$ is continuous and locally nonconstant with convex level sets. Then there exists a measurable function V_α that identifies the optimal α-forecast, i.e.,*

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

Proof: By Theorem 8 in Steinwart et al. (2014) there exists a function V_α such that for all $P \in \mathcal{P}$ it holds that $t = \alpha(P) \iff \mathbb{E}_{Y \sim P}[V_\alpha(t, Y)] = 0$. Using the same arguments as in the proof of Lemma 2, we see that

$$X = \alpha(Y \mid \mathcal{F}) \iff \mathbb{E}[V_\alpha(X, Y) \mid \mathcal{F}] = 0,$$

and an application of Lemma 1 completes the proof. □

S2 Asymmetric and state-dependent functionals

We argue that the commonly used mean functional is unsatisfactory in many situations, and that there is a need for *asymmetric* and *state-dependent* functionals.

The work of Ehm et al. (2016) attaches economic meaning to quantiles and expectiles: An agent facing the decision whether or not to invest in a certain project can identify the profit-maximizing strategy with knowledge of only a quantile or expectile of the profit distribution. Asymmetry arises e.g. in the context of tax credits for losses. If the level of tax reduction depends on profits or other time-dependent covariates, the optimal decision is a function of a state-dependent expectile. In household consumption, Andersen et al. (2008) find state-dependent risk preferences with respect to personal finances. Patton & Timmermann (2007b) argue that the Greenbook gross domestic product (GDP) forecasts are optimal with respect to an asymmetric loss function, where the level of asymmetry depends on the current growth level.

There is a vast literature arguing for the use of asymmetric loss functions (e.g., Skouras 2007, Christoffersen & Diebold 1996, 1997). For general classes of data-generating processes that allow for varying conditional means and variances but assume a constant shape of the innovation distribution, Patton & Timmermann (2007a) show that asymmetric loss leads to an optimal point forecast that is a quantile of the predictive distribution. If we allow that the shape of the innovation distribution varies, the level of the quantile may depend on a state variable.

Another source of asymmetric and state-dependent functionals is asymmetric information. Even under the symmetric mean-forecast, asymmetric information may lead to asymmetric and state-dependent functionals relative to the information set of the forecast user. As an example, consider the data-generating process

$$Y_t = Z_t^f + Z_t^u + \epsilon_t,$$

where the value of the random variable Z_t^f is known exclusively to the forecaster. The random variable Z_t^u has a mean of zero, and its value is known exclusively to the forecast user. The innovation ϵ_t has distribution F with a mean of zero, and its value is unknown to both agents. All three variables are independent. The forecaster issues the optimal mean-forecast

$$X_t = \mathbb{E}[Y_t \mid Z_t^f] = Z_t^f.$$

Under the information set of the forecast user, which is generated by Z_t^u and X_t, the point forecast X_t can be interpreted as a state-dependent quantile at level

$$\begin{aligned}\mathbb{P}(Y_t \leq X_t \mid Z_t^u, X_t) &= \mathbb{P}(Z_t^f + Z_t^u + \epsilon_t \leq Z_t^f \mid Z_t^u, X_t)\\ &= \mathbb{P}(\epsilon_t \leq -Z_t^u \mid Z_t^u)\\ &= F(-Z_t^u).\end{aligned}$$

Therefore, under asymmetric information even a standard mean-forecast may become asymmetric and state-dependent.

Finally, optimal point forecasts that are asymmetric and state-dependent can arise through transformations. For a simple example, consider a mean-forecast $X = \mathbb{E}[Y]$ for a strictly positive variable, which undergoes a logarithmic transformation and is communicated to the forecast user as $X' = \log X$. It is well known that the transformed forecast X' does not constitute an optimal mean-forecast for $Y' = \log Y$, as $\mathbb{E}[Y'] = \mathbb{E}[\log Y] < \log(\mathbb{E}[Y]) = X'$ by Jensen's inequality unless the law of Y is a point measure. The asymmetry level of the optimal mean-forecast for Y' depends on the spread of the predictive distribution, and so it can become asymmetric and state-dependent.

S3 State-dependent quantiles and expectiles

State-dependent quantile and expectile forecasts arise implicitly in essentially universal ways. Suppose that $\mathcal{L}(Y_t \mid \mathcal{F}_t)$ is continuous with a strictly positive density on its support. Irrespectively of any rationality assumptions, for any forecast X_t that is $\mathcal{F}_t$-measurable and in the support of $\mathcal{L}(Y_t|\mathcal{F}_t)$, there exists an $\mathcal{F}_t$-measurable function τ_t such that

$$X_t = q_{\tau_t}(Y_t \mid \mathcal{F}_t). \tag{S1}$$

The same argument can be applied to the expectile model, in that there exists an $\mathcal{F}_t$-measurable function υ_t such that

$$X_t = e_{\upsilon_t}(Y_t \mid \mathcal{F}_t). \tag{S2}$$

While in general τ_t and υ_t remain implicit, they can be made explicit under additional assumptions.

Specifically, let $F_{\mu,\sigma}$ denote a probability measure with cumulative distribution function (CDF)

$$F_{\mu,\sigma}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

for $x \in \mathbb{R}$, where Φ is the CDF of a fixed random variable with mean zero and variance one. Suppose that the shape of the correctly specified predictive distribution is time-invariant, so that

$$\mathcal{L}(Y_t \mid \mathcal{F}_t) = F_{\mu_t,\sigma_t}, \tag{S3}$$

where μ_t and σ_t are real-valued and strictly positive $\mathcal{F}_t$-measurable random variables, respectively. In other words, the conditional distributions $\mathcal{L}(Y_t \mid \mathcal{F}_t)$ are all members of a given location-scale family.

In the case of quantiles we suppose furthermore that Φ is absolutely continuous with a strictly positive density. If (S3) holds and X_t is in support of $\mathcal{L}(Y_t \mid \mathcal{F}_t)$, then

$$\tau_t = \Phi\left(\frac{X_t-\mu_t}{\sigma_t}\right) \tag{S4}$$

in (S1). For a particularly instructive example, suppose that the time series Y_t is generated by a stationary autoregression with mean μ, autoregressive parameters $\alpha_1, \ldots, \alpha_p$, and Gaussian innovation with variance σ^2. If the point forecast X_t is the mean of the predictive distribution under the correct autoregressive specification and order, but using incorrectly specified (e.g., estimated) parameters $\hat{\mu}$, $\hat{\alpha}_1, \ldots, \hat{\alpha}_p$, and $\hat{\sigma}^2$, then X_t is a state-dependent quantile under a specification model that is linear in the most recent observations $Y_{t-1}, \ldots, Y_{t-p}$ with a probit link function. If the autoregression is of order $p = 1$ then Y_{t-1} and X_t are affine functions of each other, and the model is linear in either of these variables.

Generally, suppose that the specification model is linear in the point forecast, so that $\tau_t = \Phi(\theta_0 + \theta_1 X_t)$ for time-independent constants θ_0 and θ_1. We can then solve (S4) for the point forecast X_t, to yield the closed form solution

$$X_t = \frac{\mu_t + \theta_0 \sigma_t}{1 - \theta_1 \sigma_t}, \tag{S5}$$

provided that $\theta_1 \sigma_t \neq 1$. This relation can be employed usefully to simulate from the linear specification model as exemplified in Section S8 below.

In the case of expectiles, if (S3) holds and X_t is in support of $\mathcal{L}(Y_t \mid \mathcal{F}_t)$ then

$$\upsilon_t = \Psi\left(\frac{X_t - \mu_t}{\sigma_t}\right)$$

in (S2), where the function Ψ can be expressed in terms of the standardized CDF Φ as

$$\Psi(x) = \frac{P(x) - x\,\Phi(x)}{2(P(x) - x\,\Phi(x)) + x}$$

for $x \in \mathbb{R}$, with $P(x) = \int_{-\infty}^{x} z \,\mathrm{d}\Phi(z)$ being the partial moment function. Interestingly, Jones (1994) showed that Ψ is also a CDF, whence every state-dependent quantile is a state-dependent expectile for a transformed conditional distribution and vice versa.

S4 Consistency of the GMM estimator

In order to establish consistency for the GMM estimator (3) in Section 3 of the main paper, we extend the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to the dynamic prediction space setting of Strähl & Ziegel (2017) and apply classical GMM theory (Hansen 1982). As noted, statements about all time points $t \in \mathbb{Z}$ are typically written without subscripts. We define $u = (x, y, s)$ and denote the identification function by $V(u, \theta)$, where $V(u, \theta) = \mathbb{1}(y \leq x) - m(s, \theta)$ in the case of quantiles and $V(u, \theta) = |\mathbb{1}(y \leq x) - m(s, \theta)|(x - y)$ in the case of expectiles.

We employ the following assumptions.

B_1 The stochastic process (U_t, W_t) is strictly stationary and ergodic.

$\mathrm{B}_{2,Q}$ The components of the instrument vector W have finite first moment.

$B_{2,E}$ The components of the vector $(X - Y)W$ have finite first moment.

B_3 The parameter space $\Theta \subseteq \mathbb{R}^p$ is compact, and the specification model $m(s, \theta)$ is continuous on Θ for all s and Borel measurable for each $\theta \in \Theta$.

B_4 The specification model is uniquely identified by the instrument vector W, i.e.,

$$\mathbb{E}[V(U, \theta)W] = 0 \iff \theta = \theta_*.$$

B_5 The weighting matrix M_T converges almost surely to a constant matrix with full rank.

Concerning the unique identification condition B_4, general insights apply, in that a specification model with p parameters calls for an instrument vector of dimension $q \geq p$. Consider for example a univariate model $m(s, \theta)$ that is strictly monotone in θ for all s. As the identification functions of quantiles and expectiles are oriented, the trivial (constant) instrument suffices for unique identification; cf. Elliott et al. (2005, Proposition 1).

For another example, let

$$m(s, \theta) = \theta' s = \sum_{i=1}^{p} \theta_i s_i$$

for $s = (s_1, \ldots, s_p)'$ with components that are pairwise uncorrelated and have positive variances. This naïve linear specification model operates under severe constraints on the domain of the parameters and the state variables, but is attractive from the perspective of interpretation. Under the quantile model the identification function satisfies $V(u, \theta) = \mathbb{1}(y \leq x) - m(s, \theta)$ and so it holds that $\mathbb{E}[V(U, \theta) \mid \mathcal{F}] = (\theta_* - \theta)' S$ and, consequently,

$$\mathbb{E}[V(U, \theta)W] = 0 \iff (\theta_* - \theta)' \, \mathbb{E}\left[\left(\sum_{i=1}^{p} S_i\right) W\right] = 0.$$

As $\mathbb{E}[S_i^2] + \sum_{j=1, j \neq i}^{k} \mathbb{E}[S_i S_j] = \mathbb{E}[S_i^2] > 0$ for $i = 1, \ldots, p$, condition B_4 is satisfied when $W = S$.

Generally, a specification model is uniquely identified if the instrument vector W generates the information set $\mathcal{F}$ and $m(S, \theta) = m(S, \theta_*)$ only if $\theta = \theta_*$. To see this, note that

$$\mathbb{E}[(V(U, \theta) - V(U, \theta_*))W] = \mathbb{E}[\, \mathbb{E}[V(U, \theta) - V(U, \theta_*) \mid \mathcal{F}] \, W] = 0$$

in concert with $\sigma(W) = \mathcal{F}$ implies $\mathbb{E}[V(U,\theta) - V(U,\theta_*)) \mid \mathcal{F}] = 0$. For quantile models it holds that $\mathbb{E}[V(U,\theta) - V(U,\theta_*) \mid \mathcal{F}] = m(S,\theta) - m(S,\theta_*)$, so $\theta = \theta_*$, which implies B_4. For expectile models $\mathbb{E}[V(U,\theta) - V(U,\theta_*) \mid \mathcal{F}] = (m(S,\theta) - m(S,\theta_*))\,\mathbb{E}[X - Y \mid \mathcal{F}]$, whence $\theta = \theta_*$ under the further condition that X is not the optimal mean-forecast.

For a first step estimator with the identity matrix as weighting matrix condition B_5 is trivially satisfied. Corollary 1 in Section S5 provides sufficient conditions for the consistency of the HAC estimator (Newey & West 1987) of the covariance matrix of the moment function, which implies B_5 for a sub-sequence of the inverse of the HAC matrix in the efficient two-step procedure.

Theorem 1 (Consistency) *If X_t is an optimal state-dependent quantile, i.e.,*

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

with some $\mathcal{F}_t$-measurable state variable S_t, conditions A_Q, B_1, $B_{2,Q}$, B_3, B_4, and B_5 guarantee existence and almost sure convergence of the GMM estimator. Analogously, if X_t is an optimal state-dependent expectile, conditions A_E, B_1, $B_{2,E}$, B_3, B_4, and B_5 yield existence and almost sure convergence.

Proof: It suffices to verify the conditions of Theorem 2.1 in Hansen (1982), namely, Assumptions 2.1–2.5, (i), (ii), and (iii). Assumption 2.1 follows from B_1 as $g(\theta) = V(u,\theta)w$ is a function of finitely many, jointly stationary and ergodic variables. Assumptions 2.2 and 2.3 are immediate from B_3, Assumption 2.4 is guaranteed by $B_{2,Q}$ or $B_{2,E}$ along with B_4 as $|g(\theta)|$ is bounded by $|w|$ in the case of quantiles and $|(x-y)w|$ in the case of expectiles, and Assumption 2.5 follows from B_5. Finally, an application of Lemma 2.1 in Hansen (1982) under $B_{2,Q}$ or $B_{2,E}$ establishes (i), B_3 yields (ii), and B_4 and B_5 guarantee (iii). □

S5 Asymptotic normality of the GMM estimator

Drawing again on well established GMM theory (Hansen 1982), we proceed to state sufficient conditions for consistent covariance estimation and asymptotic normality of the GMM estimator. As before, we consider the function $g(\theta) = V(u, \theta)w$ as a mapping from $\Theta \subseteq \mathbb{R}^p$ into $\mathbb{R}^q$. Consistency is now understood in the sense of convergence in probability.

C_1 The stochastic process (U_t, W_t) is strictly stationary with mixing coefficients α_m of order $\mathcal{O}(m^{-s})$ for some $s > 2$.

$C_{2,Q}$ There exists $\delta > 0$ such that the components of the instrument vector W have finite absolute moment of order $4 + \delta$.

$C_{2,E}$ There exists $\delta > 0$ such that the components of the vector $(X - Y)W$ have finite absolute moment of order $4 + \delta$.

C_3 The true parameter value θ_* is in the interior of Θ.

C_4 The derivative $m_{(\theta)}(s, \theta)$ exists is bounded and locally Lipschitz continuous at θ_* uniformly in s, i.e. there exists $\delta > 0$ such that $|m_{(\theta)}(s, \theta) - m_{(\theta)}(s, \theta_*)| \leq K|\theta - \theta_*|$ for all s and all θ with $|\theta - \theta_*| < \delta$.

C_5 The matrix $G = \mathbb{E}[g_{(\theta)}(\theta_*)]$ exists, is finite, and has full rank.

C_6 The weighting matrix M_T converges in probability to a constant matrix M with full rank.

C_7 The matrix $\mathbb{E}[\mathbb{1}(X \neq Y)WW']$ exists, is finite, and has full rank.

As compared to the assumption of stationarity and ergodicity in B_1, condition C_1 enforces stationarity and α-mixing, which implies ergodicity (White 2014, Proposition 3.44). The stronger condition is essential, as asymptotic normality fails generically under data generating processes with long memory (Beran 1994). Assumptions B_1 and C_1 are satisfied if the forecast and the state variable are functions of finite inputs only (e.g., based on rolling windows of data), subject to the respective stationarity and mixing conditions (cf. Giacomini & White 2006). The moment constraints in $C_{2,Q}$ and $C_{2,E}$ can be weakened at the expense of stronger mixing conditions (Hansen 1992), as generally any loosening of the regularity conditions tends to require balancing.

Theorem 2 (Asymptotic normality) *For an optimal state-dependent quantile and a consistent GMM estimator of the form in* (3)*, conditions* A_Q*,* C_1*,* $C_{2,Q}$*,* C_3*,* C_4*,* C_5*, and* C_6 *guarantee asymptotic normality with asymptotic covariance matrix*

$$\left(GM^{-1}G'\right)^{-1}\left(GM^{-1}\Sigma M^{-1}G'\right)\left(GM^{-1}G'\right)^{-1'}. \tag{S6}$$

Analogously, for an optimal state-dependent expectile and a consistent GMM estimator conditions A_E*,* C_1*,* $C_{2,E}$*,* C_3*,* C_4*,* C_5*, and* C_6 *yield the same conclusion.*

Proof: The assumption of consistency along with C_4 guarantee that the GMM estimator $\hat{\theta}_T$ in (3) satisfies Definition 3.1 of Hansen (1982). Therefore, it suffices to verify the conditions of Theorem 3.1 of Hansen (1982), namely, Assumptions 3.1–3.6, with conditions C_1, C_3, and C_4 covering Assumptions 3.1, 3.2, and 3.3, respectively. Conditions $\mathrm{C}_{2,Q}$ or $\mathrm{C}_{2,E}$ along with C_4 and C_5 imply Assumption 3.4. Lemma 3.2 in Hansen (1982), C_5, and C_6 yield Assumption 3.6.

We proceed to verify Assumption 3.5. For ease of notation, let $g_t = g_t(\theta_*)$ and let $\mathcal{I}_{t-j}$ be the σ-algebra generated by $g_{t-j}, g_{t-j-1}, \ldots$ Then $\mathbb{E}[g_t g_t']$ exists and is finite by $\mathrm{C}_{2,Q}$ or $\mathrm{C}_{2,E}$ respectively. The same conditions along with the mixing inequalities in Lemma 1.3 of Ibragimov (1962) or Theorem 14.2 of Davidson (1994), applied with $p = 2$ and $q = 4$, imply that $\mathbb{E}[g_t \mid \mathcal{I}_{t-j}] \to 0$ as $j \to \infty$ in mean square. Letting $v_j = \mathbb{E}[g_t \mid \mathcal{I}_{t-j}] - \mathbb{E}[g_t \mid \mathcal{I}_{t-j-1}]$, it remains to be shown that $\sum_{j=0}^{\infty} \mathbb{E}[v_j' v_j]^{1/2}$ is finite, which follows from the aforementioned mixing inequality in concert with the triangle and Hölder's inequalities.

We may now invoke Theorem 3.1 of Hansen (1982), which shows that $\sqrt{T}(\hat{\theta}_T - \theta_*) \to \mathcal{N}_p(0, V)$ as $T \to \infty$, where $V = (GM^{-1}G')^{-1}(GM^{-1}\Sigma M^{-1}G')(GM^{-1}G')^{-1'}$, as stated. □

In particular, Theorems 1 and 2 guarantee the consistency and asymptotic normality of the first step GMM estimator, for which the unit matrix serves as weighting matrix. We proceed to apply the latter result in order to demonstrate the asymptotic distribution (4) of the efficient two-step GMM estimator $\hat{\theta}_T$ based on a consistent first step estimator $\hat{\theta}_T^1$ and an associated heteroskedasticity and autocorrelation consistent (HAC, Newey & West 1987) estimator Σ_T of the covariance matrix Σ, with Σ_T^{-1} serving as weighting matrix. This requires condition C_7, which in the case of quantiles reduces to the standard assumption that the matrix $\mathbb{E}[WW']$ has full rank, given that A_Q implies $X \neq Y$ almost surely. In the case of expectiles C_7 generally is neither necessary nor sufficient for $\mathbb{E}[WW']$ to have full rank.

Corollary 1 (Two-step GMM with HAC covariance estimator) *For an optimal state-dependent quantile and a consistent GMM estimator $\hat{\theta}_T^1$, conditions A_Q, C_1, $C_{2,Q}$, C_3, C_4, and C_5 guarantee the consistency of the HAC estimator Σ_T that is based on $\hat{\theta}_T^1$. If furthermore condition C_7 holds true, the two-step GMM estimator $\hat{\theta}_T$ is asymptotically normal with asymptotic covariance matrix $(G\Sigma^{-1}G')^{-1}$.*

Analogously, for an optimal state-dependent expectile and a consistent GMM estimator $\hat{\theta}_T^1$, conditions A_E, C_1, $C_{2,E}$, C_3, C_4, and C_5 guarantee the consistency of the HAC estimator. If furthermore condition C_7 holds true, the two-step GMM estimator $\hat{\theta}_T$ is asymptotically normal with asymptotic covariance matrix $(G\Sigma^{-1}G')^{-1}$.

Proof: We first show that the HAC estimator Σ_T is consistent for Σ by verifying the conditions of Theorem 2 in Newey & West (1987), namely, Assumptions (i), (ii), (iii), (iv), and (v). Assumptions (i) and (ii) are guaranteed by $C_{2,Q}$ or $C_{2,E}$ along with C_4. Assumption (iii) is immediate from C_1, and to verify (iv) we apply Theorem 2. By our implementation choices, (v) is trivially satisfied. Thus, the HAC estimator is consistent for Σ. In the case of expectiles we have $\Sigma = \mathbb{E}[(\mathbb{1}(Y \leq X) - m(S, \theta_*))^2 \, (X - Y)^2 \, WW']$. If $\delta'\Sigma\delta = 0$ for some $\delta \in \mathbb{R}^q$, then $(\mathbb{1}(Y \leq X) - m(S, \theta_*))^2 (X - Y)^2 |W'\delta|^2 = 0$ almost surely, which implies $\delta = 0$ in view of condition C_7, whence Σ has full rank. In the case of quantiles a similar argument applies.

Therefore, Σ_T^{-1} is consistent for Σ^{-1}, and we may apply Theorem 2 to the efficient two-step estimator $\hat{\theta}_T$ with weighting matrix $M_T = \Sigma_T^{-1}$. Invoking (S6) we see that $\hat{\theta}_T$ is asymptotically normal with asymptotic covariance matrix $(G\Sigma^{-1}G')^{-1}$. □

S6 Conditional distributions under information rigidities

Under the data-generating process (7) in Section 4 of the main paper,

$$Y_t = \tfrac{1}{2} Y_{t-1} + \sigma_t \epsilon_t = \tfrac{1}{4} Y_{t-2} + \tfrac{1}{2} \sigma_{t-1} \epsilon_{t-1} + \sigma_t \epsilon_t.$$

Conditioning on $\mathcal{I}_{t-2}$ yields $\mathbb{E}[Y_t \mid \mathcal{I}_{t-2}] = \frac{1}{4} Y_{t-2}$. Furthermore, $\mathcal{L}(Y_t \mid \mathcal{I}_{t-2})$ is symmetric, so $q_{1/2}(Y_t \mid \mathcal{I}_{t-2}) = \mathbb{E}[Y_t \mid \mathcal{I}_{t-2}] = \frac{1}{4} Y_{t-2}$.

S7 State-independent forecasts under different information sets: Additional results

In this section, we provide additional results under the simulation setting in Section 4.1 of the main paper.

First, we assess the finite-sample relevance of the asymptotic distribution (4) of the GMM estimator, as discussed in Sections 3.2 and S5. Specifically, consider the linear quantile specification model with the point forecast X_t of eq. (8) as state variable. The true parameter values in this setting are $\theta_0 = 0.10$ and $\theta_1 = 0.25$. Figure S1 illustrates the empirical distribution of the GMM estimates based on sample paths of size $T = 100$, 250, and 1000, respectively. There is good agreement with the large sample approximation.

Figure S2 is the equivalent of Figure 3 in the main paper, except that our tests are now based on expectile models. Figure S3 shows results from the same experiment, but with the quantile- and expectile-based tests now using the same high-dimensional instrument vector as for the spline approach. While we discourage doing so in small samples, the quantile- and expectile-based tests show the desired asymptotic behaviour, whereas the spline test does not.

Similarly, Figure S4 is the equivalent of Figure 4 in the main paper, but our tests are now based on quantile models. State-dependent optimality allows for a broad class of forecasting behavior, and therefore the linear test can be expected to detect suboptimal forecasts at

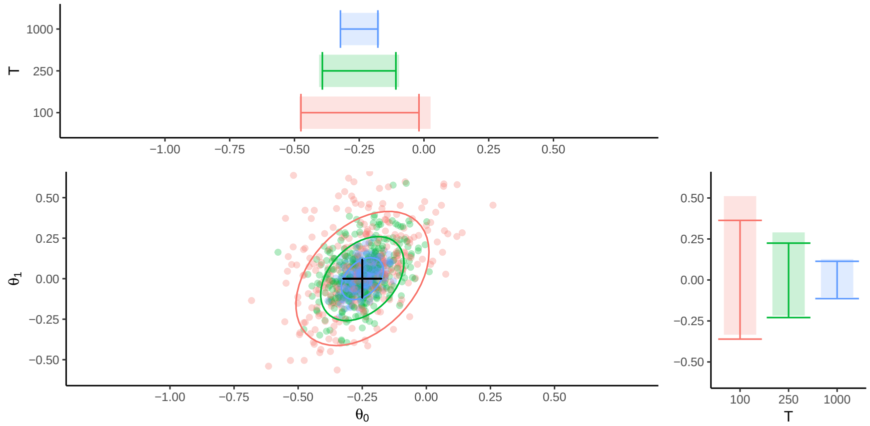

Figure S1: Empirical distribution of the GMM estimate under the linear quantile specification model based on sample paths of size $T = 100, 250$, and 1000 from (7) and (8). The scatterplot shows estimates of $\theta = (\theta_0, \theta_1)'$ along with the respective 90% ellipsoids from the large sample approximation. The true parameter values are at the center of the cross. The boxes at top and right range from the 5th to the 95th percentile of the estimates, as compared to the large sample approximation (bars).

a constant level less often than the more specific constant test. Perhaps surprisingly, in this setting the additional degree of freedom in the state-dependent linear test does not lead to a decrease in finite-sample power. Finally, Figure S5 shows the results from the same experiment, but with the quantile- and expectile-based tests using the same high-dimensional instrument vector as for the spline approach. While at small sample sizes, the quantile- and expectile-based tests are undersized and less powerful than under the three-dimensional instrument vector, they exhibit much higher size-adjusted power than the spline-based test.

S8 State-dependent quantile forecasts in terms of X_t

In Section 4.2 of the main paper we construct optimal state-dependent quantile-forecasts based on break, linear, and periodic specification models in the state variable Y_{t-1}, and we investigate the size and power of overidentifying restrictions tests for forecast optimality under the respective models. Table S1 in this supplement considers the case where the

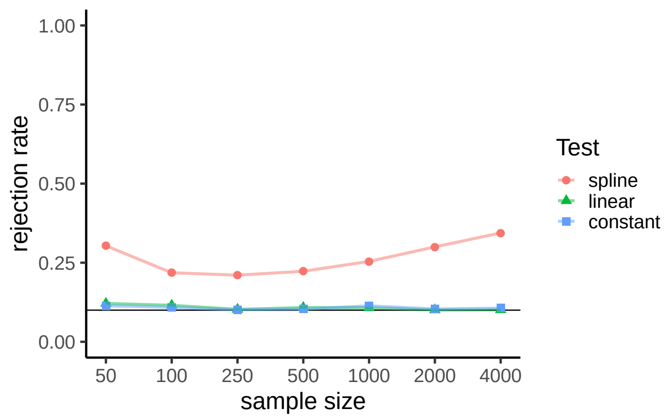

Figure S2: Same as Figure 3 in the main paper, but with our tests now based on expectile models: Size of optimality tests for the one-step ahead forecast. The horizontal line is at the nominal level of 0.10.

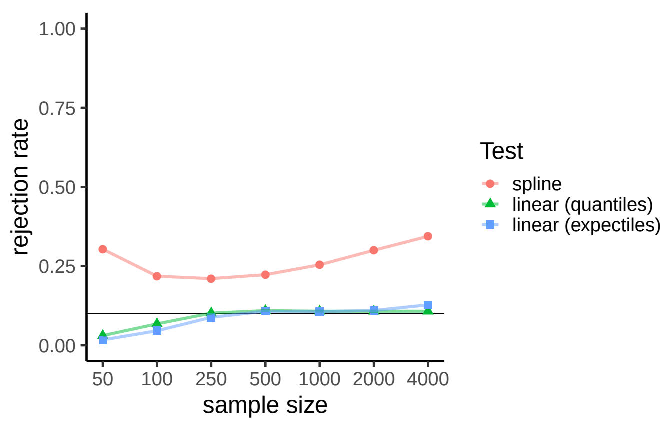

Figure S3: Same as Figures 3 and S2, but with the quantile- and expectile-based tests now using the same seven-dimensional instrument vector as for the spline test.

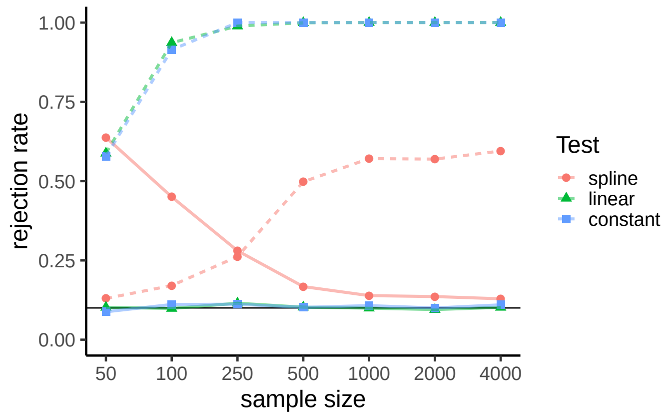

Figure S4: Same as Figure 4 in the main paper, but with our tests now based on quantile models: Size and size-adjusted power of optimality tests for the two-step ahead forecast. The horizontal line is at the nominal level of 0.10. The solid lines represent size for tests with properly lagged instruments. The dashed lines represent size-adjusted power for tests with nonlagged instruments.

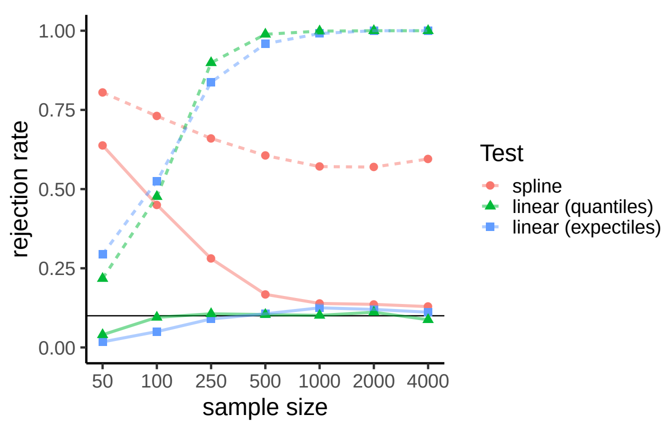

Figure S5: Same as Figures 4 and S4, but with the quantile- and expectile-based linear tests now using the same seven-dimensional instrument vector as for the spline test.

Table S1: Rejection rates of optimality tests based on quantile specification models in the state variable X_t, when the true model is linear. 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. Settings where the null hypothesis is naturally satisfied are marked in gray.

	hypothesized model					
sample	$W_t = (1, Y_{t-1}, X_t)'$			$W_t = (1, Y_{t-1}, Y_{t-2})'$		
size	linear	break	periodic	linear	break	periodic
$T = 100$	0.10	0.18	0.17	0.10	0.09	0.09
$T = 250$	0.11	0.34	0.43	0.10	0.10	0.10
$T = 1000$	0.12	0.74	0.94	0.10	0.10	0.11

state variable in the linear quantile specification model is the point forecast X_t, rather than the most recent outcome, Y_{t-1}, while all other facets of the data generating process, the GMM estimator and the test are retained. Optimal state-dependent forecasts at a level that depends on X_t are generally defined implicitly only. However, under the linear specification model simple closed form expressions are available, as detailed in equation (S5). Again, the optimality test is well calibrated and depends crucially on the choice of the instrument vector.

S9 State-dependent forecasts under different specification models: Results for tests based on expectiles

As noted, Section 4.2 of the main paper considers optimal state-dependent quantile-forecasts based on break, linear, and periodic specification models in the state variable Y_{t-1}. In this supplementary section, we retain the data generating process, and the hypothesized specification models are still the break, linear, and periodic models, but now expressed in terms of expectiles rather than quantiles.

Table S2: Rejection rates of expectile and quantile optimality tests in the setting of Section 4.2 with instrument vector $W_t = (1, Y_{t-1}, X_t)'$. Settings where the null hypothesis is satisfied are marked in gray.

	true model	hypothesized model					
sample size	quantile	expectile			quantile		
		linear	break	periodic	linear	break	periodic
$T = 100$	linear	0.09	0.30	0.38	0.10	0.23	0.36
	break	0.35	0.12	0.07	0.28	0.11	0.08
	periodic	0.73	0.37	0.06	0.65	0.30	0.07
$T = 250$	linear	0.10	0.63	0.83	0.11	0.42	0.87
	break	0.65	0.12	0.16	0.52	0.10	0.15
	periodic	0.99	0.71	0.07	0.96	0.58	0.09
$T = 1000$	linear	0.10	0.98	0.99	0.11	0.86	1.00
	break	0.99	0.10	0.68	0.96	0.10	0.44
	periodic	1.00	1.00	0.09	1.00	0.99	0.10

Results for both the expectile- and quantile-based tests of forecast optimality with instrument vector $W_t = (1, Y_{t-1}, X_t)'$ are presented in Table S2. If the type of the model is correctly specified, the quantile-forecasts do not get rejected by the expectile-based tests; in fact, the expectile tests are nearly as calibrated as and slightly more powerful than the quantile tests.

As shown in Section S3, for any given distribution, the quantile at a given level is simply an expectile at another level. For state-dependent forecasts this implies the existence of a specification model $m'(s)$ such that $q_{m(s)}(Y \mid \mathcal{F}) = e_{m'(s)}(Y \mid \mathcal{F})$. The transformation from m to m' depends on the conditional predictive distribution, but remains the same within location-scale families (Yao & Tong 1996, Proposition 1), such as in our simulation setting, where all conditional distributions are Gaussian. Therefore, the transformation is static, and for the break model, which specifies two levels only, the choice of quantiles versus expectiles affects the implementation of the test, but not the model itself. Jones (1994) notes that the expectiles for the standard normal distribution are closely approximated by the quantiles of a normal distribution with standard deviation $\frac{2}{3}$. Thus, the linear and periodic quantile forecasts are approximately linear and periodic expectile forecasts with scaled parameters. Generally, increased flexibility in the specification model makes the distinction between models in terms of quantiles and expectiles harder and potentially superfluous.

References

Andersen, S., Harrison, G. W., Lau, M. I. & Rutström, E. (2008), 'Lost in state space: Are preferences stable?', International Economic Review **49**, 1091–1112.

Beran, J. (1994), Statistics for Long-Memory Processes, Chapman & Hall, Boca Raton.

Billingsley, P. (1995), Probability and Measure, 3rd edn, John Wiley & Sons, New York.

Christoffersen, P. F. & Diebold, F. X. (1996), 'Further results on forecasting and model selection under asymmetric loss', Journal of Applied Econometrics **11**, 561–571.

Christoffersen, P. F. & Diebold, F. X. (1997), 'Optimal prediction under asymmetric loss', Econometric Theory **13**, 808–817.

Davidson, J. (1994), Stochastic Limit Theory: An Introduction for Econometricians, Oxford University Press, Oxford.

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.

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.

Hansen, B. E. (1992), 'Consistent covariance matrix estimation for dependent heterogeneous processes', Econometrica **60**, 967–972.

Hansen, L. P. (1982), 'Large sample properties of generalized method of moments estimators', Econometrica **50**, 1029–1054.

Ibragimov, I. A. (1962), 'Some limit theorems for stationary processes', Theory of Probability & Its Applications **7**, 349–382.

Jones, M. C. (1994), 'Expectiles and M-quantiles are quantiles', Statistics & Probability Letters **20**, 149–153.

Newey, W. K. & West, K. D. (1987), 'A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix', Econometrica **55**, 703–708.

Ok, E. A. (2007), Real Analysis with Economic Applications, Princeton University Press, Princeton.

Osband, K. H. (1985), Providing Incentives for Better Cost Forecasting, PhD thesis, University of California, Berkeley.

Patton, A. J. & Timmermann, A. (2007a), 'Properties of optimal forecasts under asymmetric loss and non-linearity', Journal of Econometrics **140**, 884–918.

Patton, A. J. & Timmermann, A. (2007b), 'Testing forecast optimality under unknown loss', Journal of the American Statistical Association **102**, 1172–1184.

Skouras, S. (2007), 'Decisionmetrics: A decision-based approach to econometric modelling', Journal of Econometrics **137**, 414–440.

Steinwart, I., Pasin, C., Williamson, R. & Zhang, S. (2014), Elicitation and identification of properties, in M. F. Balcan & C. Szepesvári, eds, 'Conference on Learning Theory 2014', Vol. 35 of Proceedings of Machine Learning Research, pp. 482–526.

Strähl, C. & Ziegel, J. (2017), 'Cross-calibration of probabilistic forecasts', Electronic Journal of Statistics **11**, 608–639.

White, H. (2014), Asymptotic Theory for Econometricians, Academic Press, London.

Yao, X. & Tong, H. (1996), 'Asymmetric least squares estimation: A nonparametric approach', Nonparametric Statistics **6**, 273–292.