Interpretation of point forecasts with unknown directive

Patrick Schmidt
(joint with Tilmann Gneiting and Matthias Katzfuss)

Heidelberg Institute for Theoretical Studies and Goethe University Frankfurt

ES/ASSA, Philadelphia, 2018

Motivation

- Forecasts should be probabilistic!

Motivation

- Forecasts should be probabilistic!

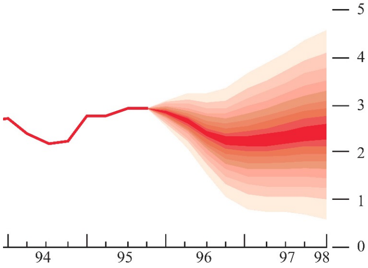


Figure: Source: Bank of England. Inflation report, February 1996.

Motivation

- Forecasts should be probabilistic!

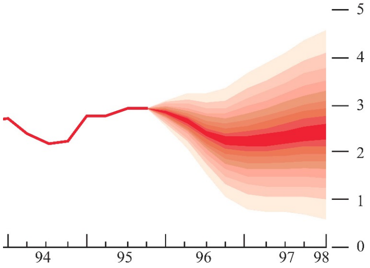


Figure: Source: Bank of England. Inflation report, February 1996.

- Testing rationality ✓

Motivation

- Forecasts should be probabilistic!

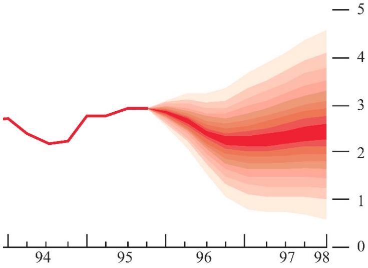


Figure: Source: Bank of England. Inflation report, February 1996.

- Testing rationality ✓
- Comparing forecasts ✓

Motivation

- Forecasts should be probabilistic!

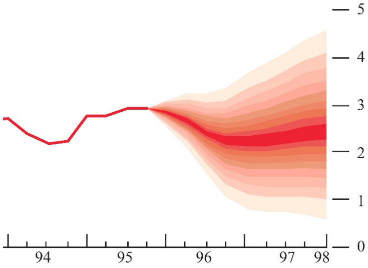


Figure: Source: Bank of England. Inflation report, February 1996.

- Testing rationality ✓
- Comparing forecasts ✓
- Best response / economic action ✓

Motivation

- However, point forecasts are still common
  - Historic reasons
  - Convenience
  - Complex computer models

Motivation

- However, point forecasts are still common
  - Historic reasons
  - Convenience
  - Complex computer models

- Point forecasts without directive contain ambiguous information
  (Manski, 2016; Elliott and Timmermann, 2008; Engelberg et al., 2009; Gneiting, 2011)

Motivation

- However, point forecasts are still common
  - Historic reasons
  - Convenience
  - Complex computer models
- Point forecasts without directive contain ambiguous information
  (Manski, 2016; Elliott and Timmermann, 2008; Engelberg et al., 2009; Gneiting, 2011)
- How can we conceptualize and estimate the directive?

Motivation

- However, point forecasts are still common
  - Historic reasons
  - Convenience
  - Complex computer models

- Point forecasts without directive contain ambiguous information
  (Manski, 2016; Elliott and Timmermann, 2008; Engelberg et al., 2009; Gneiting, 2011)

- How can we conceptualize and estimate the directive?
  - Elliott et al. (2005) and Patton and Timmermann (2007) propose to estimate the underlying loss function

Summary

- In this paper, we consider the situation of an unknown directive

Summary

- In this paper, we consider the situation of an unknown directive

- We propose a setting in which point forecasts are derived as functionals (i.e., point summaries) from predictive distributions

Summary

- In this paper, we consider the situation of an unknown directive
- We propose a setting in which point forecasts are derived as functionals (i.e., point summaries) from predictive distributions
- We show how to identify the functional based on forecast and realization alone

Summary

- In this paper, we consider the situation of an unknown directive
- We propose a setting in which point forecasts are derived as functionals (i.e., point summaries) from predictive distributions
- We show how to identify the functional based on forecast and realization alone
- We propose a GMM-estimator for state-dependent quantile (expectile) forecasts

Summary

- In this paper, we consider the situation of an unknown directive
- We propose a setting in which point forecasts are derived as functionals (i.e., point summaries) from predictive distributions
- We show how to identify the functional based on forecast and realization alone
- We propose a GMM-estimator for state-dependent quantile (expectile) forecasts
- We find that the Federal Reserves' GDP forecasts are quantiles that depend on the current growth level

Outline

Setting

- A forecaster predicts the random variable Y_t based on information $\mathcal{F}_t$

Setting

- A forecaster predicts the random variable Y_t based on information $\mathcal{F}_t$

- Common interpretation of a point forecast X_t:

$$X_t = \mathbb{E}[Y_t | \mathcal{F}_t]$$

Setting

- A forecaster predicts the random variable Y_t based on information $\mathcal{F}_t$
- Common interpretation of a point forecast X_t:

$$X_t = \mathbb{E}[Y_t|\mathcal{F}_t]$$

- Let $\alpha : \mathcal{P} \mapsto \mathbb{R}$ be some functional (Horowitz and Manski, 2006)

Setting

- A forecaster predicts the random variable Y_t based on information $\mathcal{F}_t$
- Common interpretation of a point forecast X_t:

$$X_t = \mathbb{E}[Y_t|\mathcal{F}_t]$$

- Let $\alpha : \mathcal{P} \mapsto \mathbb{R}$ be some functional (Horowitz and Manski, 2006)

Definition

We call X_t an *optimal α-forecast*, if it holds that

$$X_t = \alpha(Y_t|\mathcal{F}_t).$$

Definition

We call X_t an *optimal* α*-forecast*, if it holds that

$$X_t = \alpha(Y_t|\mathcal{F}_t).$$

Definition

We call X_t an *optimal α-forecast*, if it holds that

$$X_t = \alpha(Y_t | \mathcal{F}_t).$$

- The functional α and the conditional distributions $\mathcal{L}(Y_t | \mathcal{F}_t)$ are unobservable

Definition

We call X_t an *optimal α-forecast*, if it holds that

$$X_t = \alpha(Y_t|\mathcal{F}_t).$$

- The functional α and the conditional distributions $\mathcal{L}(Y_t|\mathcal{F}_t)$ are unobservable
- We want to do inference on α

Outline

Identification

- For an optimal mean-forecast there exist identifying moment conditions:

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

Identification

- For an optimal mean-forecast there exist identifying moment conditions:

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

- We show that, for well-behaved α, there exists an *identification function* V_α, such that

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

(Proof: Steinwart et al. (2014) + Gneiting et al. (2007))

Loss functions

- Previous work (Elliott et al., 2005; Patton and Timmermann, 2007) defined optimal point forecasts with loss functions $L(x, y)$ by
$$X = \arg\min_{x \in \mathbb{R}} \mathbb{E}[L(x, Y)|\mathcal{F}]. \tag{1}$$
- Defines a functional
- There exist many loss functions (Gneiting, 2011; Ehm et al., 2016) for one functional
$\Rightarrow$ It is impossible to identify the loss function.

Outline

Setting

Identification

Parametric estimation

Application

MCS

Discussion

State-dependent quantile forecasts

- The τ-quantile functional $q_\tau(\mathbb{P})$ is unique solution x to $\mathbb{P}((-\infty, x]) = \tau$

State-dependent quantile forecasts

- The τ-quantile functional $q_\tau(\mathbb{P})$ is unique solution x to $\mathbb{P}((-\infty, x]) = \tau$

- Identifying moment condition

$$X = q_\tau(Y|\mathcal{F}) \iff \mathbb{E}[\mathbb{1}(Y \leq x) - \tau|\mathcal{F}] = 0.$$

State-dependent quantile forecasts

- The τ-quantile functional $q_\tau(\mathbb{P})$ is unique solution x to $\mathbb{P}((-\infty, x]) = \tau$

- Identifying moment condition

$$X = q_\tau(Y|\mathcal{F}) \iff \mathbb{E}[\mathbb{1}(Y \leq x) - \tau|\mathcal{F}] = 0.$$

- We assume an optimal quantile-forecast with levels described by the specification model $m(z, \theta_0)$,

$$X = q_{m(Z,\theta_0)}(Y|\mathcal{F}).$$

State-dependent quantile forecasts

- The τ-quantile functional $q_\tau(\mathbb{P})$ is unique solution x to $\mathbb{P}((-\infty, x]) = \tau$
- Identifying moment condition

$$X = q_\tau(Y|\mathcal{F}) \iff \mathbb{E}[\mathbb{1}(Y \leq x) - \tau|\mathcal{F}] = 0.$$

- We assume an optimal quantile-forecast with levels described by the specification model $m(z, \theta_0)$,

$$X = q_{m(Z,\theta_0)}(Y|\mathcal{F}).$$

specification name	model	Θ
constant	$m(z,\theta) = \theta$	$(0,1)$
break	$m(z,\theta) = \mathbb{1}(z > t_b)\theta_1 + \mathbb{1}(z \leq t_b)\theta_2$	$(0,1)^2$
linear	$m(z,\theta) = \Psi(\theta_1 + z \cdot \theta_2)$	$[a,b]^2$
periodic	$m(z,\theta) = \Psi(\theta_1 + \theta_2 \sin(2\pi z/\theta_3))$	$[a,b]^3$

GMM-Estimation

- $X = q_{m(Z,\theta_0)}(Y|\mathcal{F})$ implies
$$\mathbb{E}[(\mathbb{1}(Y \leq X) - m(Z,\theta_0)) \cdot W] = 0.$$

- Empirical mean is
$g_T(\theta) := \frac{1}{T}\sum_{t=1}^{T}(\mathbb{1}(y_t \leq x_t) - m(z_t,\theta)) \cdot w_t.$

GMM-Estimation

- $X = q_{m(Z,\theta_0)}(Y|\mathcal{F})$ implies
$$\mathbb{E}[(\mathbb{1}(Y \leq X) - m(Z,\theta_0)) \cdot W] = 0.$$

- Empirical mean is
$g_T(\theta) := \frac{1}{T}\sum_{t=1}^{T}(\mathbb{1}(y_t \leq x_t) - m(z_t,\theta)) \cdot w_t.$

- The standard GMM-estimator is obtained by
$$\hat{\theta}_T = \underset{\theta\in\Theta}{\arg\min}\, g_T(\theta)'\hat{S}_T^{-1} g_T(\theta).$$

GMM-Estimation

- $X = q_{m(Z,\theta_0)}(Y|\mathcal{F})$ implies
$$\mathbb{E}[(\mathbb{1}(Y \leq X) - m(Z,\theta_0)) \cdot W] = 0.$$
- Empirical mean is
$g_T(\theta) := \frac{1}{T}\sum_{t=1}^{T}(\mathbb{1}(y_t \leq x_t) - m(z_t,\theta)) \cdot w_t.$
- The standard GMM-estimator is obtained by
$$\hat{\theta}_T = \arg\min_{\theta\in\Theta} g_T(\theta)'\hat{S}_T^{-1}g_T(\theta).$$

If m is continuous, there exists an $\mathcal{F}_t$-measurable instrumental variable W_t such that
$$\hat{\theta}_T \xrightarrow{P} \theta_0.$$

Testing optimality

- Test of overidentifying restrictions (Hansen, 1982): If the forecast is optimal, it holds that

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

where $J_T(\theta) := Tg_T(\theta)'\hat{S}_T(\theta)^{-1}g_T(\theta)$.

Testing optimality

- Test of overidentifying restrictions (Hansen, 1982): If the forecast is optimal, it holds that

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

where $J_T(\theta) := T g_T(\theta)' \hat{S}_T(\theta)^{-1} g_T(\theta)$.

- If X is optimal with respect to $\mathcal{F}$, than it is also optimal with respect to any smaller information set.

- We can test hypothesis H_0:

There exists a $\theta \in \Theta$ such that $X = q_{m(z,\theta)}(Y|\mathcal{F})$ with $\sigma(w) \subset \mathcal{F}$.

Outline

Application: GDP forecast

- Consider the Greenbook GDP growth forecasts of the Federal Reserve
  - quarterly real GDP growth rate forecasts over the period 1968 to 2011 ($T = 172$ observations)
  - standard tests of optimality based on the mean functional reject the optimality of the forecast (Patton and Timmermann, 2007)

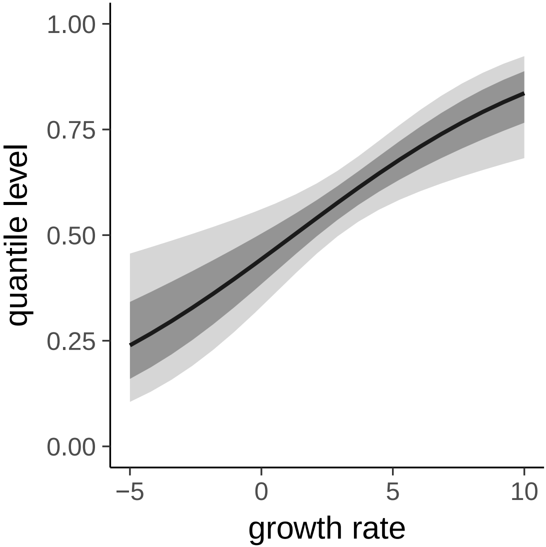
Application: GDP forecast

- Consider the Greenbook GDP growth forecasts of the Federal Reserve
  - quarterly real GDP growth rate forecasts over the period 1968 to 2011 ($T = 172$ observations)
  - standard tests of optimality based on the mean functional reject the optimality of the forecast (Patton and Timmermann, 2007)

- We interpret the forecasts as a state-dependent quantile

$$m(z_t, \theta) = m(x_t, \theta) = \Psi(\theta_1 + x_t \cdot \theta_2).$$

Results



Outline

MCS

- Data-generating process:

$$\begin{aligned} Y_t &= \tfrac{1}{2} Y_{t-1} + \sigma_t \epsilon_t \quad \text{for } t = 1, 2, \ldots, T, \\ \sigma_t^2 &= .1 + .8\sigma_{t-1}^2 + .1\sigma_{t-1}^2 \epsilon_{t-1}^2, \\ \epsilon_t &\overset{iid}{\sim} \mathcal{N}(0, 1). \end{aligned}$$

- Generate on each data set optimal forecasts and different optimality tests

Constant quantile forecast

- A fully informed 1-step ahead forecast

$$x_t = \frac{1}{2} Y_{t-1} - \frac{1}{4}\sigma_t .$$

Constant quantile forecast

- A fully informed 1-step ahead forecast

$$x_t = \frac{1}{2}Y_{t-1} - \frac{1}{4}\sigma_t.$$

- A 2-step ahead forecast

$$x_t = \frac{1}{4}Y_{t-2} + 1.15\sigma_{t-1}^2 + .1$$

Constant quantile forecast

- A fully informed 1-step ahead forecast

$$x_t = \tfrac{1}{2} Y_{t-1} - \tfrac{1}{4}\sigma_t.$$

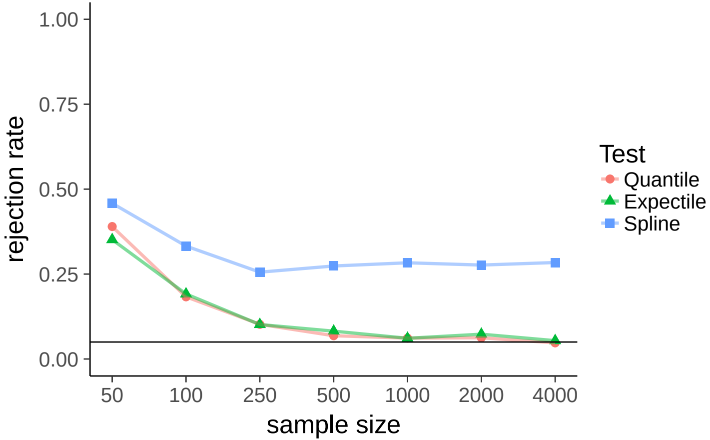
- A 2-step ahead forecast

$$x_t = \tfrac{1}{4} Y_{t-2} + 1.15\sigma_{t-1}^2 + .1$$

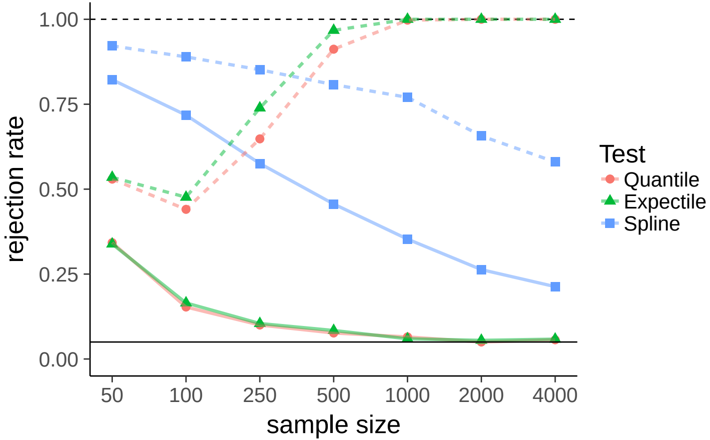
- We apply optimality tests of linear specification models for
  - quantiles,
  - expectiles,
  - and the spline test of Patton and Timmermann (2007)

  with normal and with lagged instruments.

Results for 1-step ahead forecast



Results for 2-step ahead forecast



Different state-dependent forecasts

- We generated three optimal state-dependent forecasts
  - a quantile forecast that depends linearly on the current time series value (*linear*)
  - one that is subject to periodic deviations (*periodic*)
  - one that is exposed to a break at half the sample size (*break*)

Different state-dependent forecasts

- We generated three optimal state-dependent forecasts
  - a quantile forecast that depends linearly on the current time series value (*linear*)
  - one that is subject to periodic deviations (*periodic*)
  - one that is exposed to a break at half the sample size (*break*)

- To each forecast we apply a test of optimality based on the linear, periodic and break specification model.

Results: rejection rates

$T = 1000$

forecast model	test model		
	linear	periodic	break
linear	0.059	1.000	1.000
periodic	1.000	0.057	1.000
break	1.000	1.000	0.059

Results: rejection rates

$T = 1000$

forecast model	test model		
	linear	periodic	break
linear	0.059	1.000	1.000
periodic	1.000	0.057	1.000
break	1.000	1.000	0.059

$T = 100$

forecast model	test model		
	linear	periodic	break
linear	0.092	0.643	0.784
periodic	0.844	0.076	0.619
break	0.673	0.373	0.087

Outline

Discussion

- Functionals are identifiable with point forecasts and observations alone

Discussion

- Functionals are identifiable with point forecasts and observations alone

- State-dependent quantiles/expectiles can be used to interpret point forecasts coherently

Discussion

- Functionals are identifiable with point forecasts and observations alone
- State-dependent quantiles/expectiles can be used to interpret point forecasts coherently
- Applications possible in
  - comparison of point and probability forecasts
  - backtesting of risk measures
  - creating density forecasts from multiple point forecasts

Connection to Mincer-Zarnowitz

References I

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

Elliott, G., I. Komunjer, and A. Timmermann (2005): "Estimation and testing of forecast rationality under flexible loss," *The Review of Economic Studies*, 72, 1107–1125.

Elliott, G. and A. Timmermann (2008): "Economic Forecasting," *Journal of Economic Literature*, 46, 3–56.

Engelberg, J., C. F. Manski, and J. Williams (2009): "Comparing the point predictions and subjective probability distributions of professional forecasters," *Journal of Business & Economic Statistics*, 27, 30–41.

Gneiting, T. (2011): "Making and evaluating point forecasts," *Journal of the American Statistical Association*, 106, 746–762.

Gneiting, T., F. Balabdaoui, and A. E. Raftery (2007): "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69, 243–268.

Hansen, L. P. (1982): "Large sample properties of Generalized Method of Moments estimators," *Econometrica*, 50, 1029–1054.

References II

Horowitz, J. L. and C. F. Manski (2006): "Identification and estimation of statistical functionals using incomplete data," *Journal of Econometrics*, 132, 445–459.

Manski, C. F. (2016): "Interpreting Point Predictions: Some Logical Issues," *Foundations and Trends in Accounting*, 10, 238–261.

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

Steinwart, I., C. Pasin, R. Williamson, and S. Zhang (2014): "Elicitation and Identification of Properties," in *Proceedings of The 27th Conference on Learning Theory*, ed. by M. F. Balcan, V. Feldman, and C. Szepesvári, vol. 35 of *Proceedings of Machine Learning Research*, 482–526.

Connection to Mincer-Zarnowitz test

- Procedure of Mincer-Zarnowitz optimality test
  1. OLS regression $Y = \beta_0 + \beta_1 X + u$
  2. Test H_0: $\beta_0 = 0 \wedge \beta_1 = 1$

- Mincer-Zarnowitz test is equivalent to
  1. Assume optimal forecast

  $$X = \beta_0 + \beta_1 \mathbb{E}[Y|X]$$

  2. Derive identification function

  $$\mathbb{E}[(\beta_0 + \beta_1 X - Y)|\mathcal{F}] = 0$$

  3. Apply GMM-estimator with instruments $W = (1, X)$
  4. Test H_0: $\beta_0 = 0 \wedge \beta_1 = 1$

discussion