Selecting Models with Judgment

Simone Manganelli
European Central Bank

2nd "Forecasting at Central Banks" conference

Bank of England and King's College London, 15 November, 2018

Question

Start from a judgmental forecast.

How can it be statistically improved?

Outline

The frequentist statistical decision rule

Model selection under misspecification

Empirical illustration

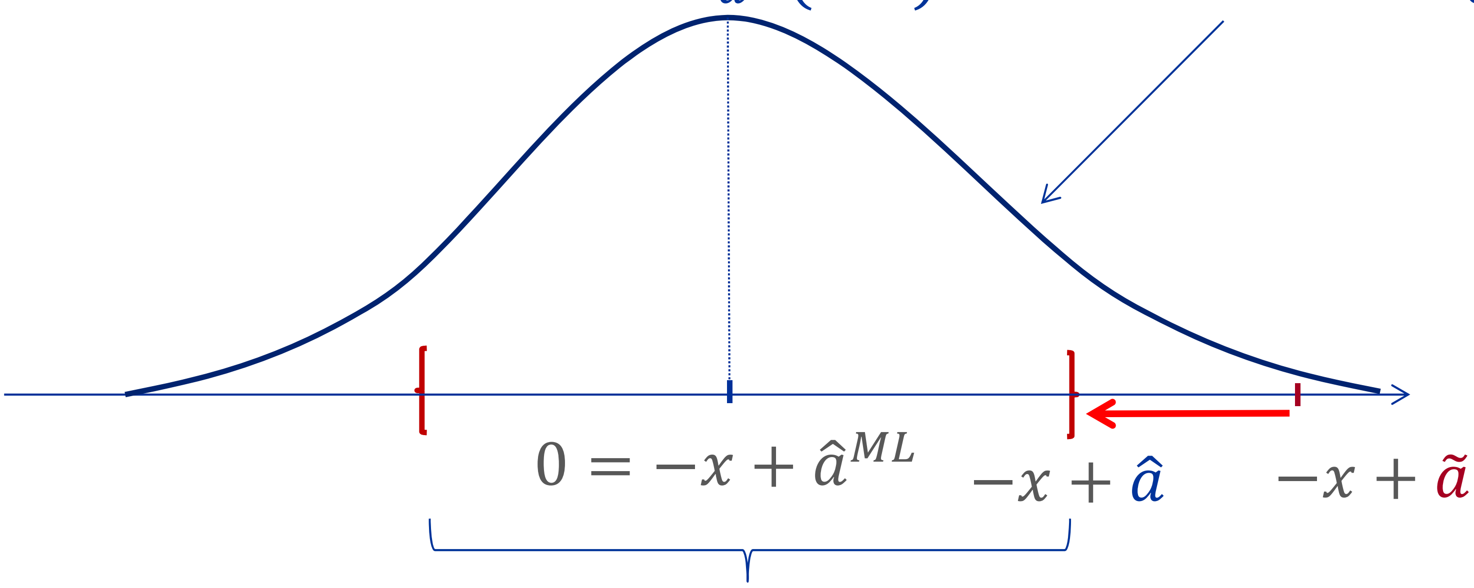
- Preliminaries
- Results

The frequentist decision rule with judgment

1) $X \sim N(\theta, 1)$ 2) $L(\theta, a) = 0.5(\theta - a)^2$
3) Judgment: $\{\tilde{a}, \alpha\}$ 4) $X = x$

FOC: H_0: $\nabla_a L(\theta, \tilde{a}) = -\theta + \tilde{a} = 0$

Distribution of $\nabla_a L(\hat{\theta}, \tilde{a}) = -X + \tilde{a}$ under $H_0: \theta = \tilde{a}$



The frequentist decision rule with judgment

1) $X \sim N(\theta, 1)$

The frequentist decision rule with judgment

1) $X \sim N(\theta, 1)$

2) $L(\theta, a) = 0.5(\theta - a)^2$

The frequentist decision rule with judgment

1) $X \sim N(\theta, 1)$ 2) $L(\theta, a) = 0.5(\theta - a)^2$
3) Judgment: $\{\tilde{a}, \alpha\}$

The frequentist decision rule with judgment

1) $X \sim N(\theta, 1)$

2) $L(\theta, a) = 0.5(\theta - a)^2$

3) Judgment: $\{\tilde{a}, \alpha\}$

4) $X = x$

The frequentist decision rule with judgment

1) $X \sim N(\theta, 1)$	2) $L(\theta, a) = 0.5(\theta - a)^2$
3) Judgment: $\{\tilde{a}, \alpha\}$	4) $X = x$

FOC: H_0: $\nabla_a L(\theta, \tilde{a}) = -\theta + \tilde{a} = 0$

The frequentist decision rule with judgment

1) $\mathrm{X} \sim N(\theta, 1)$ 2) $\mathrm{L}(\theta, a) = 0.5(\theta - a)^2$

3) Judgment: $\{\tilde{a}, \alpha\}$ 4) $\mathrm{X} = x$

FOC: H_0: $\nabla_a \mathrm{L}(\theta, \tilde{a}) = -\theta + \tilde{a} = 0$

Test statistic: $-\mathrm{X} + \tilde{a}$

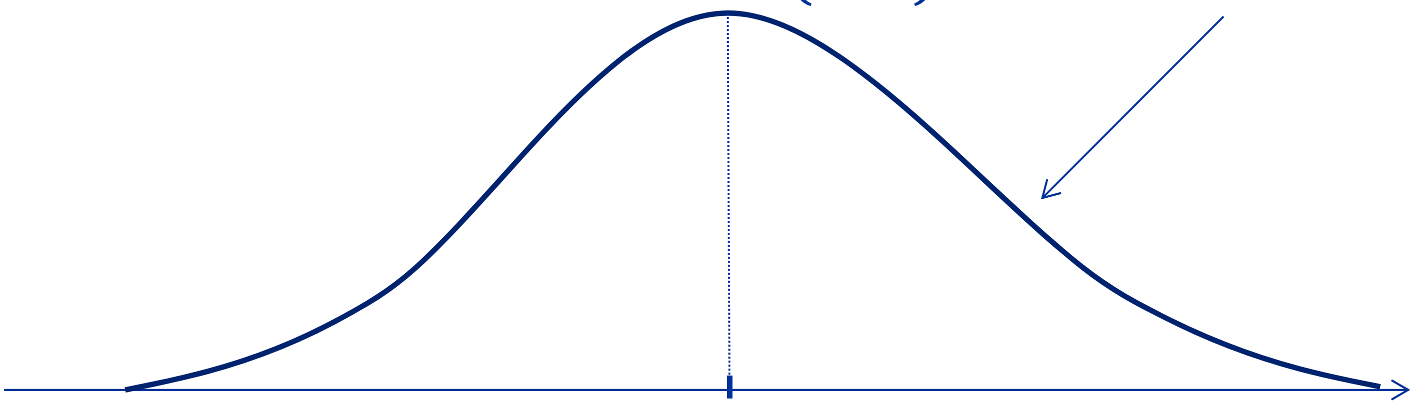
The frequentist decision rule with judgment

1) $X \sim N(\theta, 1)$
2) $L(\theta, a) = 0.5(\theta - a)^2$
3) Judgment: $\{\tilde{a}, \alpha\}$
4) $X = x$

FOC: H_0: $\nabla_a L(\theta, \tilde{a}) = -\theta + \tilde{a} = 0$

Test statistic: $-X + \tilde{a}$

Distribution of $\nabla_a L(\hat{\theta}, \tilde{a}) = -X + \tilde{a}$ under $H_0: \theta = \tilde{a}$

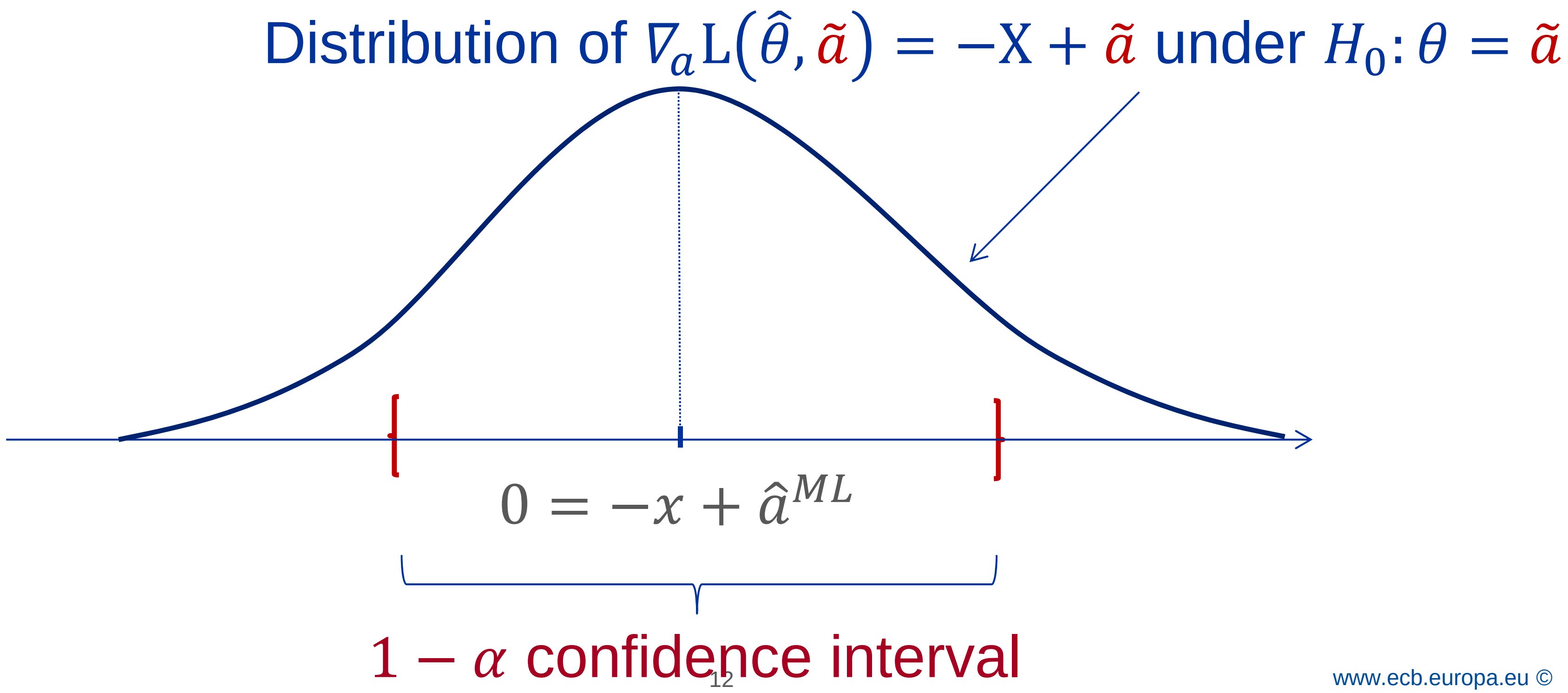


$0 = -x + \hat{a}^{ML}$

The frequentist decision rule with judgment

1) $X \sim N(\theta, 1)$ 2) $L(\theta, a) = 0.5(\theta - a)^2$
3) Judgment: $\{\tilde{a}, \alpha\}$ 4) $X = x$

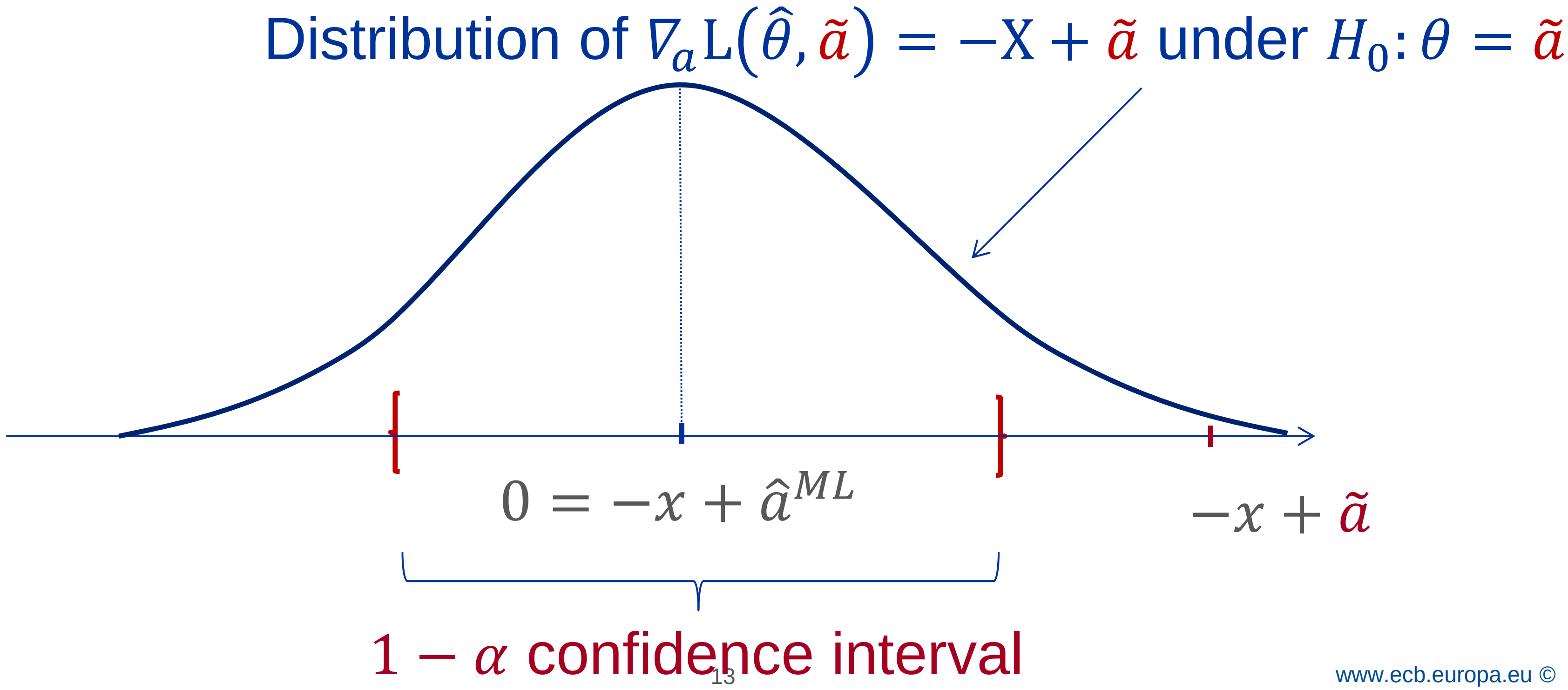
FOC: H_0: $\nabla_a L(\theta, \tilde{a}) = -\theta + \tilde{a} = 0$
Test statistic: $-X + \tilde{a}$



The frequentist decision rule with judgment

1) $X \sim N(\theta, 1)$ 2) $L(\theta, a) = 0.5(\theta - a)^2$
3) Judgment: $\{\tilde{a}, \alpha\}$ 4) $X = x$

FOC: H_0: $\nabla_a L(\theta, \tilde{a}) = -\theta + \tilde{a} = 0$
Test statistic: $-X + \tilde{a}$

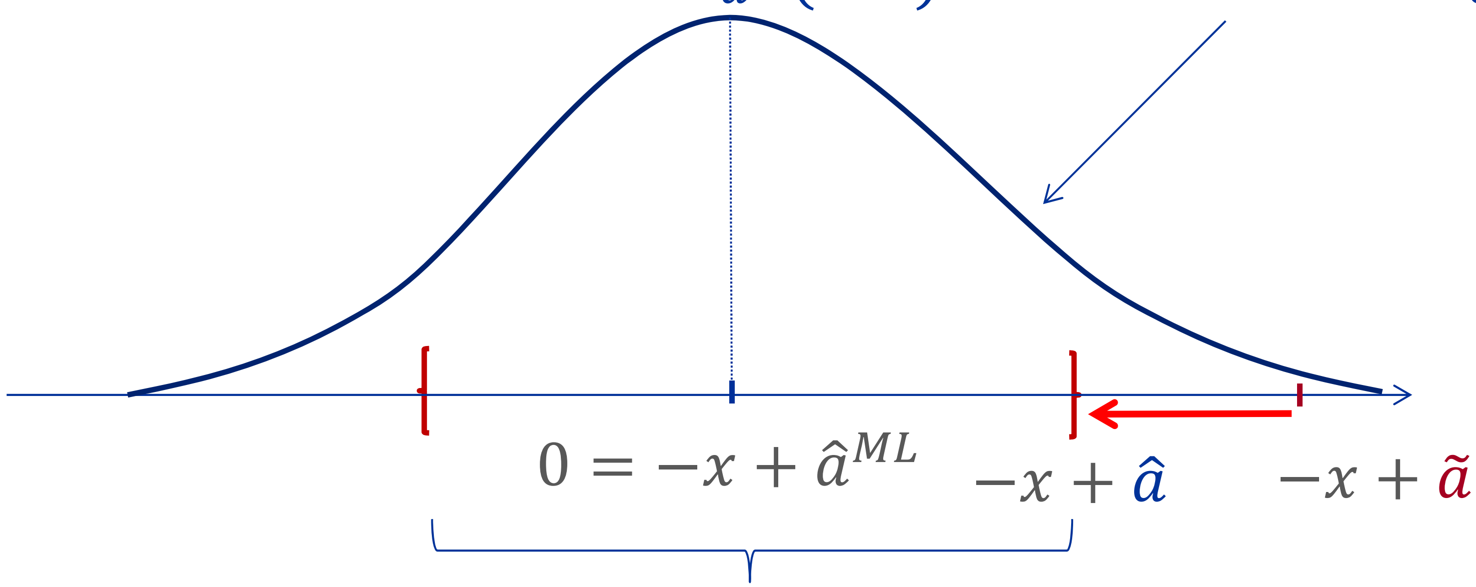


The frequentist decision rule with judgment

1) $X \sim N(\theta, 1)$	2) $L(\theta, a) = 0.5(\theta - a)^2$
3) Judgment: $\{\tilde{a}, \alpha\}$	4) $X = x$

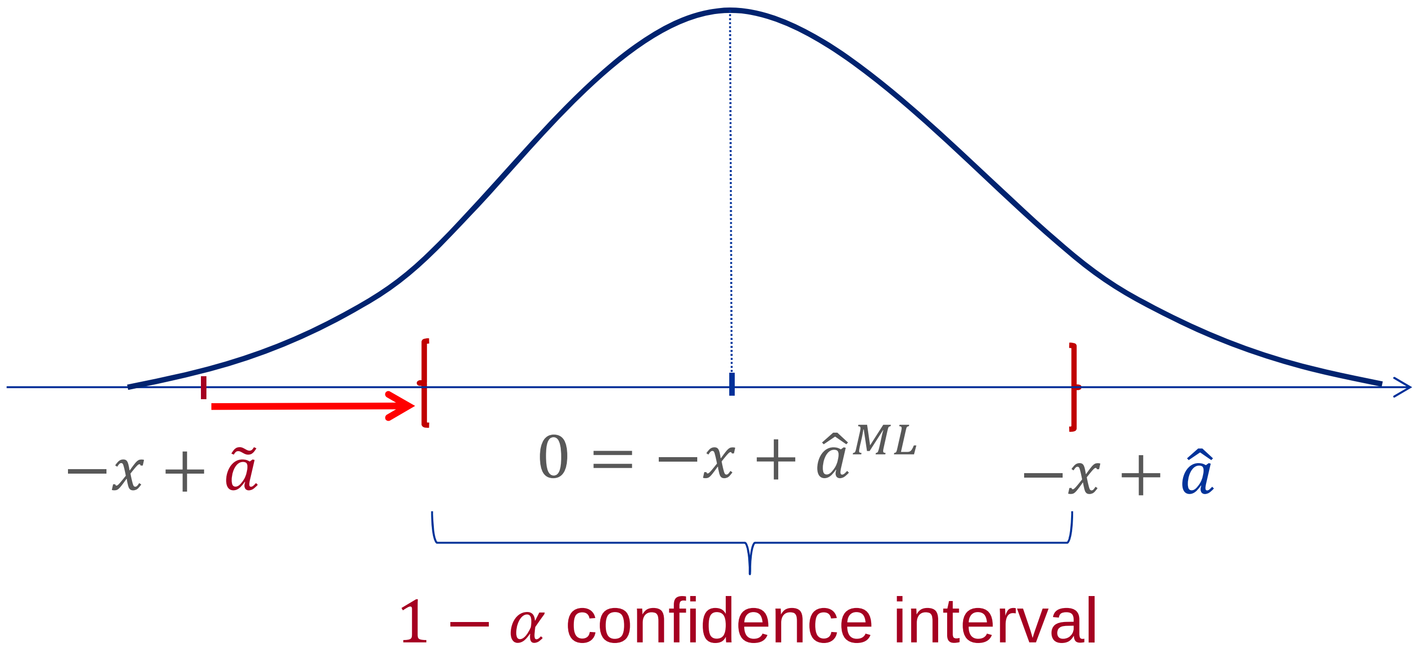
FOC: H_0: $\nabla_a L(\theta, \tilde{a}) = -\theta + \tilde{a} = 0$

Distribution of $\nabla_a L(\hat{\theta}, \tilde{a}) = -X + \tilde{a}$ under $H_0: \theta = \tilde{a}$

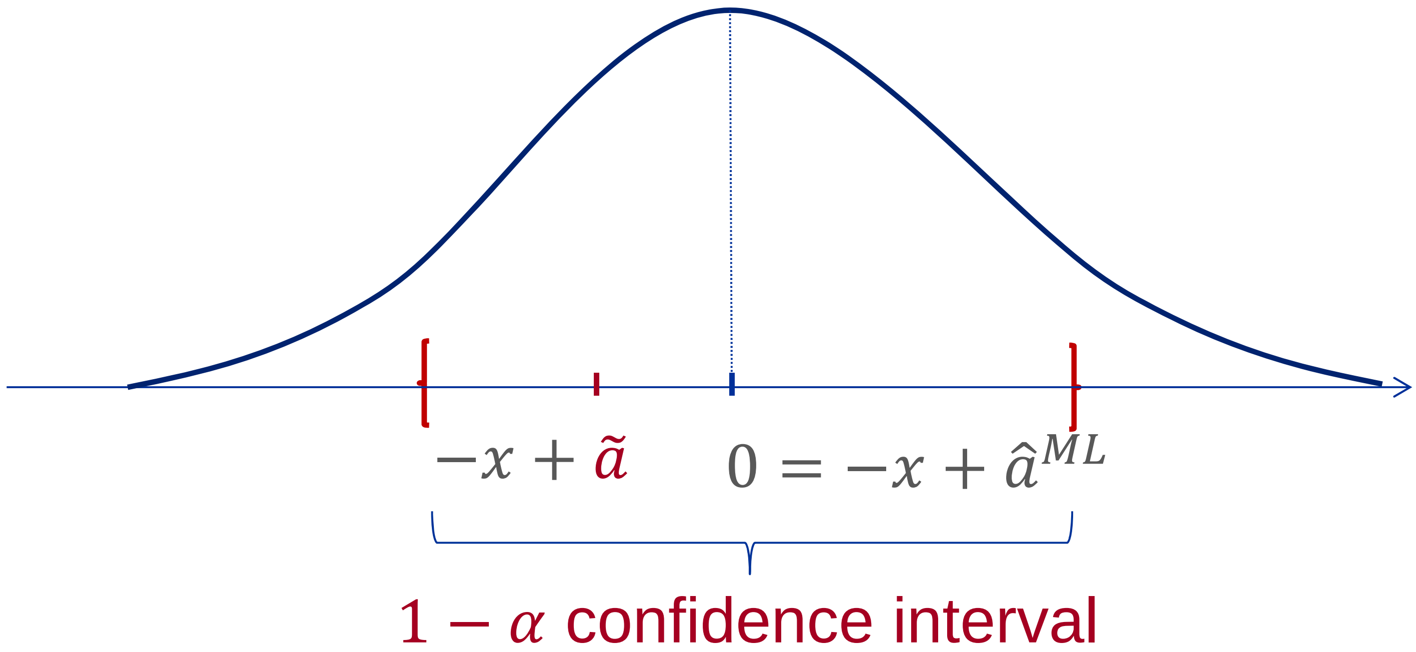


$1 - \alpha$ confidence interval

The frequentist decision rule with judgment



The frequentist decision rule with judgment



The frequentist decision

Frequentist decision rule:

$$\delta(X) = (x + c_{1-\alpha}) \cdot I(-X + \tilde{a} > c_{1-\alpha}) + \\ + (x + c_{\alpha}) \cdot I(-X + \tilde{a} < c_{\alpha}) + \\ + \tilde{a} \cdot I(c_{\alpha} \leq -X + \tilde{a} \leq c_{1-\alpha})$$

Economic interpretation:

$$P_{\theta}\left[L\left(\theta, \delta(X)\right) > L(\theta, \tilde{a})\right] \leq \alpha$$

Outline

The frequentist statistical decision rule

Model selection under misspecification

Empirical illustration

- Preliminaries
- Results

Assumptions so far:

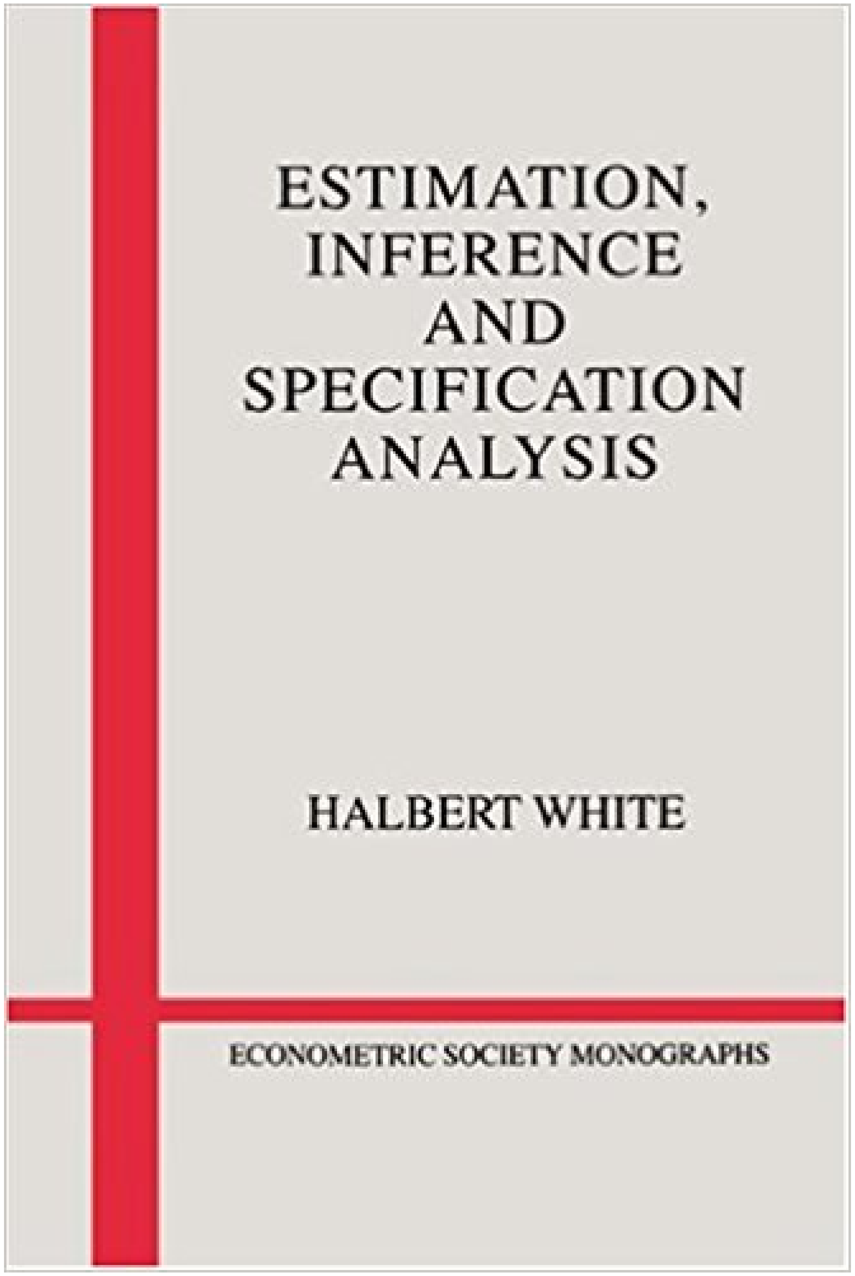
1) Quadratic loss function
   Easy to generalise to convex loss functions

2) A single observation normally distributed
   Easy to generalise, appealing to asymptotic approximations

3) Correctly specified model
   This is hard…

Halbert L. White, Jr., 1950-2012

Theory for misspecified models



Quasi-maximum likelihood estimation

$$\max_{\theta} n^{-1} \sum_{t=1}^{n} \log f_t(X^t, \theta)$$

$$B^{*-\frac{1}{2}} A^* \sqrt{n}(\hat{\theta}(X^n) - \theta^*) \xrightarrow{A} N(0, I_p)$$

Quasi-maximum likelihood estimation

$$\max_{\theta} n^{-1} \sum_{t=1}^{n} \log f_t(X^t, \theta)$$

$$B^{*-\frac{1}{2}} A^* \sqrt{n}(\hat{\theta}(X^n) - \theta^*) \xrightarrow{A} N(0, I_p)$$

Econometrics boils down to the choice of $f_t(X^t, \theta)$.

Special cases:

- Macroeconometrics (VAR, DSGE, …)
- Financial econometrics (GARCH, CAViaR, …)
- Machine learning (neural nets, random forests, …)

Quasi-maximum likelihood estimation

$$\max_{\theta} n^{-1} \sum_{t=1}^{n} \log f_t(X^t, \theta)$$

$$B^{*-\frac{1}{2}} A^{*} \sqrt{n} (\hat{\theta}(X^n) - \theta^{*}) \xrightarrow{A} N(0, I_p)$$

The asymptotic VC matrix is consistently estimated by:

$$\hat{A}^{-1} \hat{B} \hat{A}^{-1}$$

$$\hat{A} = n^{-1} \sum_{t=1}^{n} \nabla^2 \log f_t\, (X^t, \hat{\theta}(X^n))$$

$$\hat{B} = n^{-1} \sum_{t=1}^{n} \nabla \log f_t\, (X^t, \hat{\theta}(X^n)) \nabla' \log f_t\, (X^t, \hat{\theta}(X^n))$$

Judgment and asymptotics

The decision problem for inflation forecasting:

$$\min_{a} L(\theta, a) = 0.5(x_{n,h}(\theta) - a)^2$$

Judgment and asymptotics

The decision problem for inflation forecasting:

$$\min_a L(\theta, a) = 0.5(x_{n,h}(\theta) - a)^2$$

Null hypothesis that the judgment is optimal:

$$H_0: \nabla L(\theta^*, \tilde{a}) = -x_{n,h}(\theta^*) + \tilde{a} = 0$$

Use classical Wald, LR or LM tests.

Back to the simple case

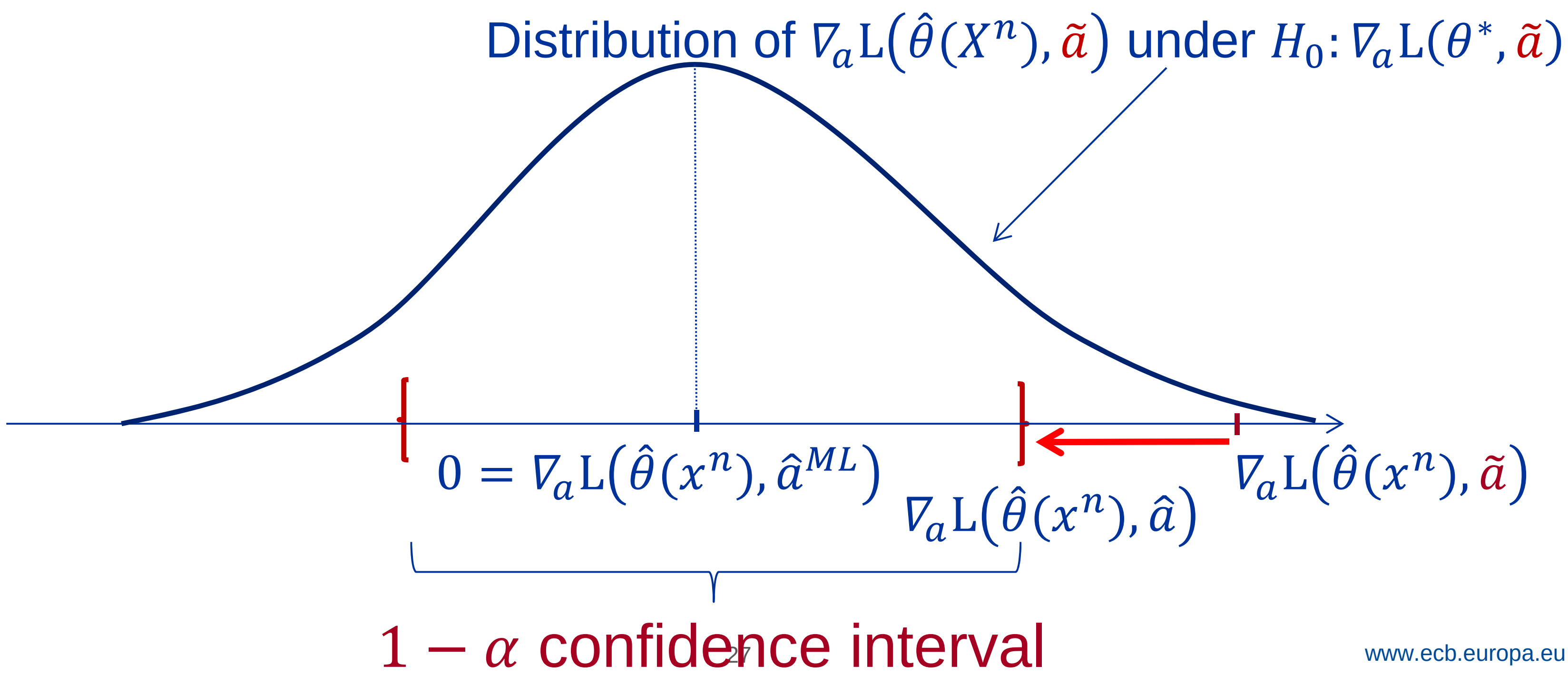
1) $x_{n,h}(\hat{\theta}(X^n)) \sim N\left(x_{n,h}(\theta^*), \sigma^2\right)$

2) $\mathrm{L}(\theta^*, a) = 0.5(x_{n,h}(\theta^*) - a)^2$

3) Judgment: $\{\tilde{a}, \alpha\}$

4) $X^n = x^n$

FOC: H_0: $\nabla_a \mathrm{L}(\theta^*, \tilde{a}) = -x_{n,h}(\theta^*) + \tilde{a} = 0$



Model selection

Select the least misspecified model.

Model selection

Select the least misspecified model.

Minimise the Kullback-Leibler distance:

$$\Pi(a^0 : \delta^m) = n^{-1} \sum_{t=1}^{n} \left(L({\delta_t}^m(x^n)\ ) - L({a_t}^0) \right) \geq 0$$

Model selection

Select the least misspecified model.

Minimise the Kullback-Leibler distance:

$$\Pi(a^0 : \delta^m) = n^{-1} \sum_{t=1}^{n} (L({\delta_t}^m(x^n)\ ) - L({a_t}^0)) \geq 0$$

Asymptotic approximation:

$$\hat{L}^m = n^{-1} \sum_{t=1}^{n} 0.5({\delta_t}^m(x^n) - x_{t,h})^2$$

Model selection

Select the least misspecified model.

Minimise the Kullback-Leibler distance:

$$\Pi(a^0 : \delta^m) = n^{-1} \sum_{t=1}^{n} (L({\delta_t}^m(x^n)\,) - L({a_t}^0)) \geq 0$$

Asymptotic approximation:

$$\hat{L}^m = n^{-1} \sum_{t=1}^{n} 0.5({\delta_t}^m(x^n) - x_{t,h})^2$$

Best model:

$$m^+ = \arg\min_m \hat{L}^m$$

Outline

The frequentist statistical decision rule

Model selection under misspecification

Empirical illustration

- **Preliminaries**
- Results

Ingredients for forecasting

1) Judgment
2) Loss function
3) Data
4) Functional form for the asymptotic approximation

Ingredients for forecasting

1) Judgment
2) Loss function
3) Data
4) Functional form for the asymptotic approximation

Judgment – Inflation forecast

Trusting that the ECB will deliver on its mandate:

$$\tilde{a} \equiv 24^{-1} E(\sum_{i=1}^{24} \pi_{n+i} \,|X^n) = 1.9\%$$

Judgment – Confidence level

$\alpha = 10\%$

Ingredients for forecasting

1) Judgment
2) Loss function
3) Data
4) Functional form for the asymptotic approximation

Quadratic loss function

$$L(a) = 0.5E((\pi_{n,24} - a)^2 | X^n)$$

where

$$\pi_{n,24} \equiv 24^{-1} \sum_{i=1}^{24} \pi_{n+i}$$

Ingredients for forecasting

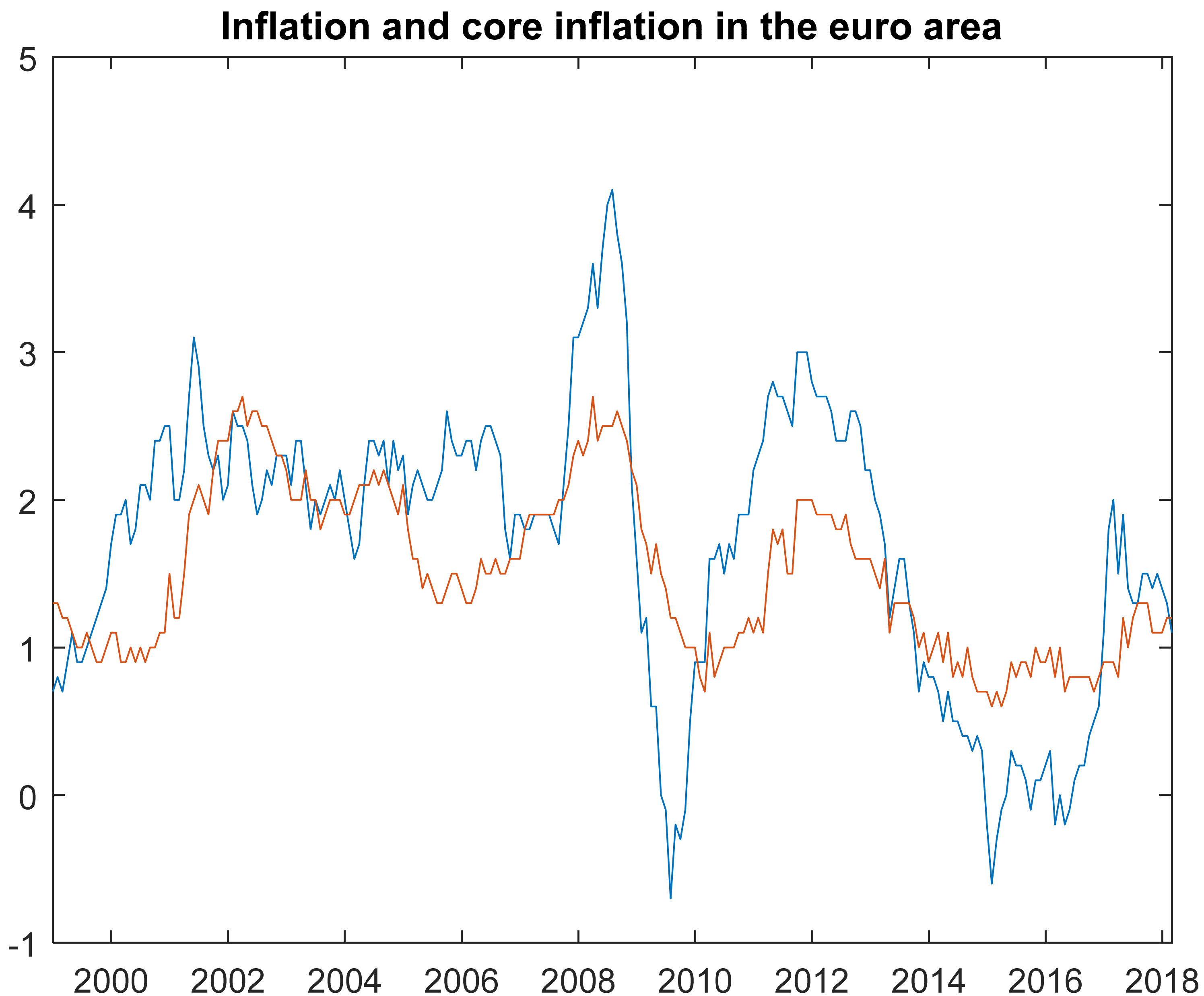
1) Judgment
2) Loss function
3) Data
4) Functional form for the asymptotic approximation

Data

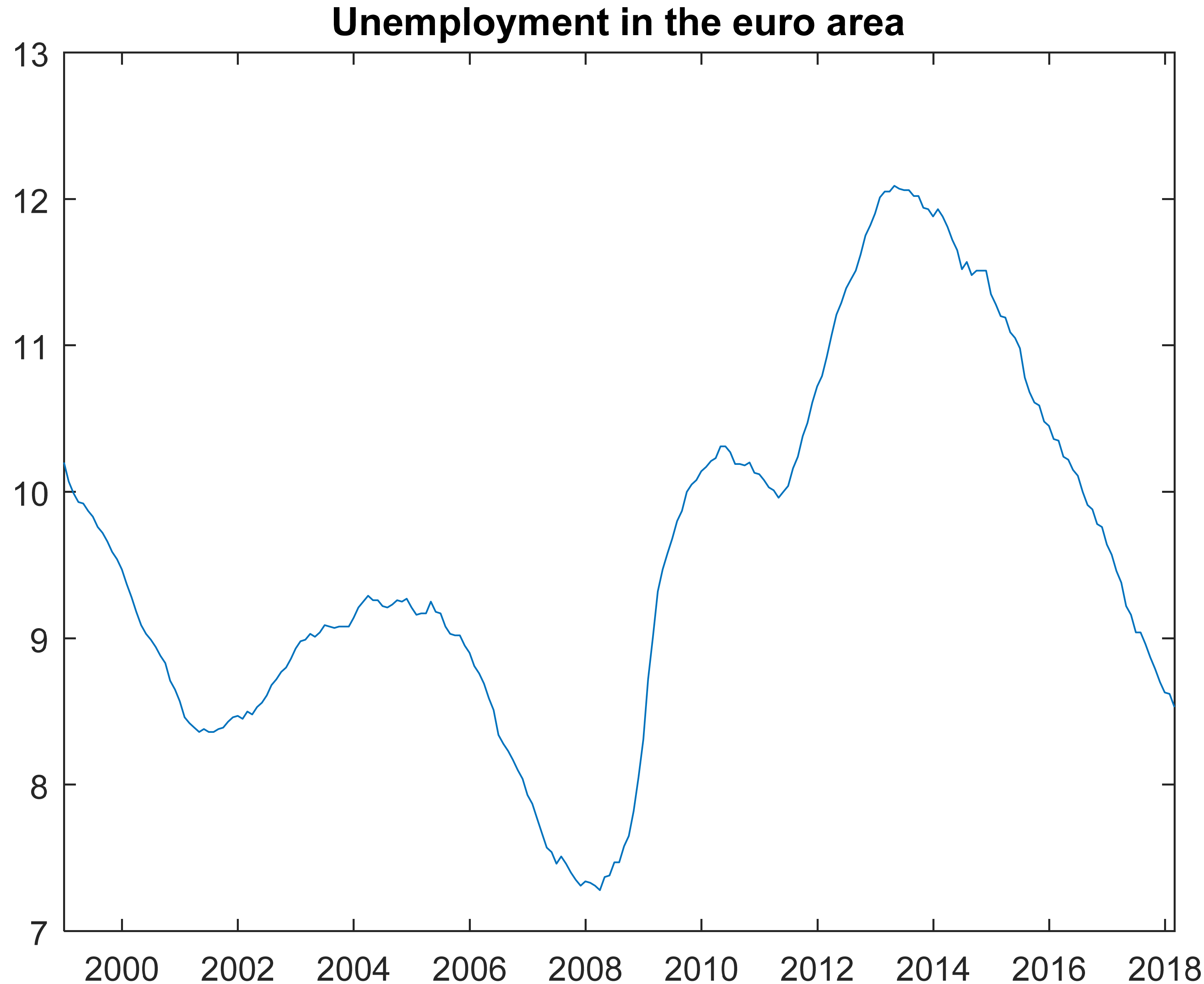
1) HICP - Overall index, Annual rate of change, Eurostat
2) HICP - All-items excluding energy and unprocessed food, Annual rate of change, Eurostat
3) Unemployment, Eurostat
4) Industrial Production Index, Eurostat

Source: https://sdw.ecb.europa.eu

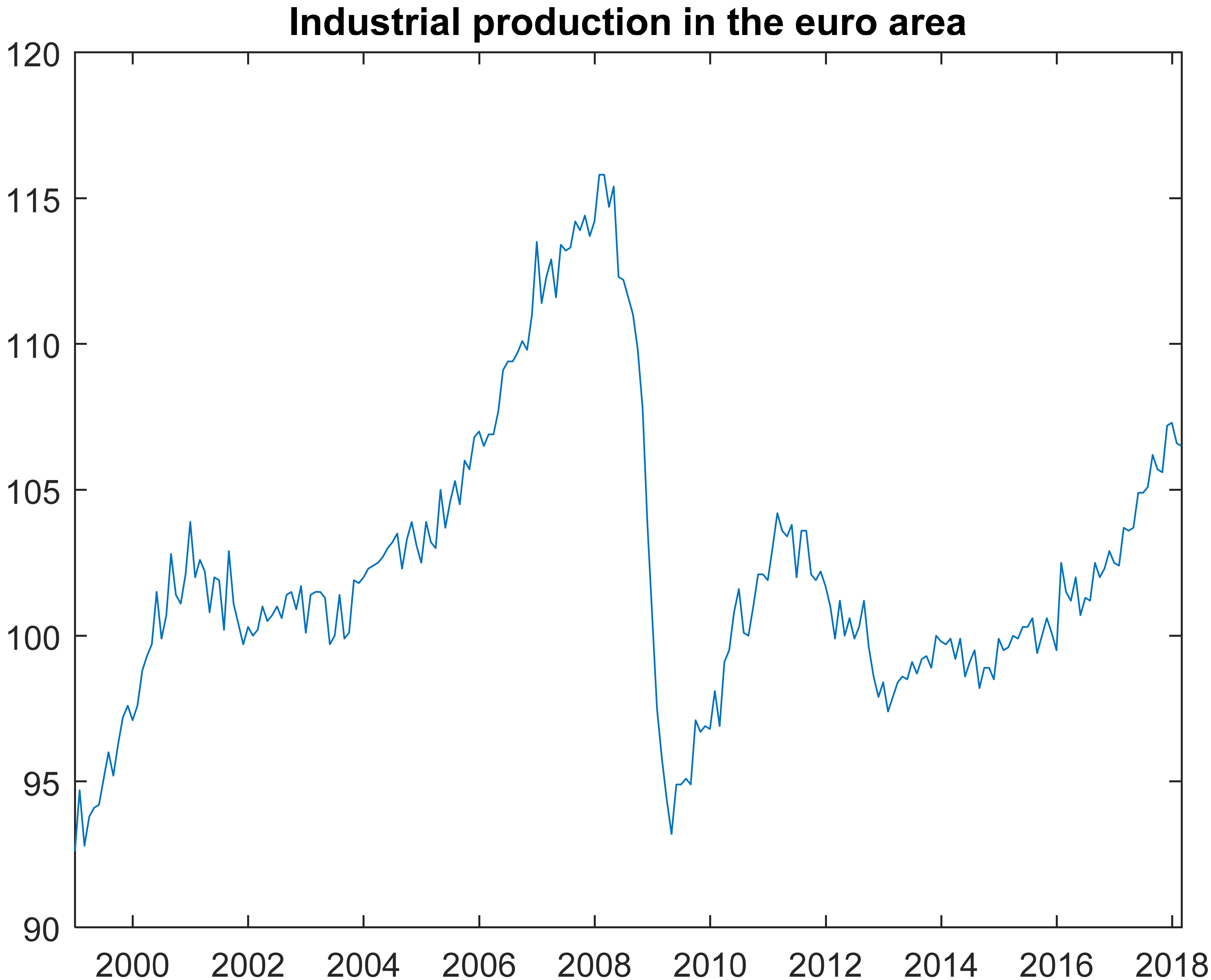
Annualised inflation



Unemployment



Industrial production



Ingredients for forecasting

1) Judgment
2) Loss function
3) Data
4) Functional form for the asymptotic approximation

VAR model

$$X_t = \upsilon + A_1 X_{t-1} + \cdots + A_p X_{t-p} + u_t$$
$$u_t \sim iid(0, \Sigma)$$

with zero restrictions:

→ ratio of parameters to data not too large

Choice of $f_t(X^t, \theta)$:

$$f_t(X^t, \theta) = \exp(-0.5 u_t{}' \Sigma^{-1} u_t)$$

Example

	c	1	2	3	4	5	6
HICP	1	1 0 1 0	0 0 0 1	0 0 0 0	0 0 0 0	0 0 0 0	1 0 0 0
HICPX	0	1 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 1 0 0
UN	0	1 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
IP	1	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
	2	4	2	0	0	0	2

	7	8	9	10	11	12	
	0 0 0 0	0 0 0 0	0 1 0 0	0 0 0 0	0 0 0 0	0 0 1 1	8
	0 0 0 0	0 0 0 0	1 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	3
	0 0 0 0	0 0 0 0	0 0 1 0	0 0 0 0	0 0 0 0	0 0 0 0	2
	0 0 0 0	0 0 0 1	0 0 1 0	0 0 0 1	0 0 0 0	0 0 0 0	5
	0	1	4	1	0	2	18

Outline

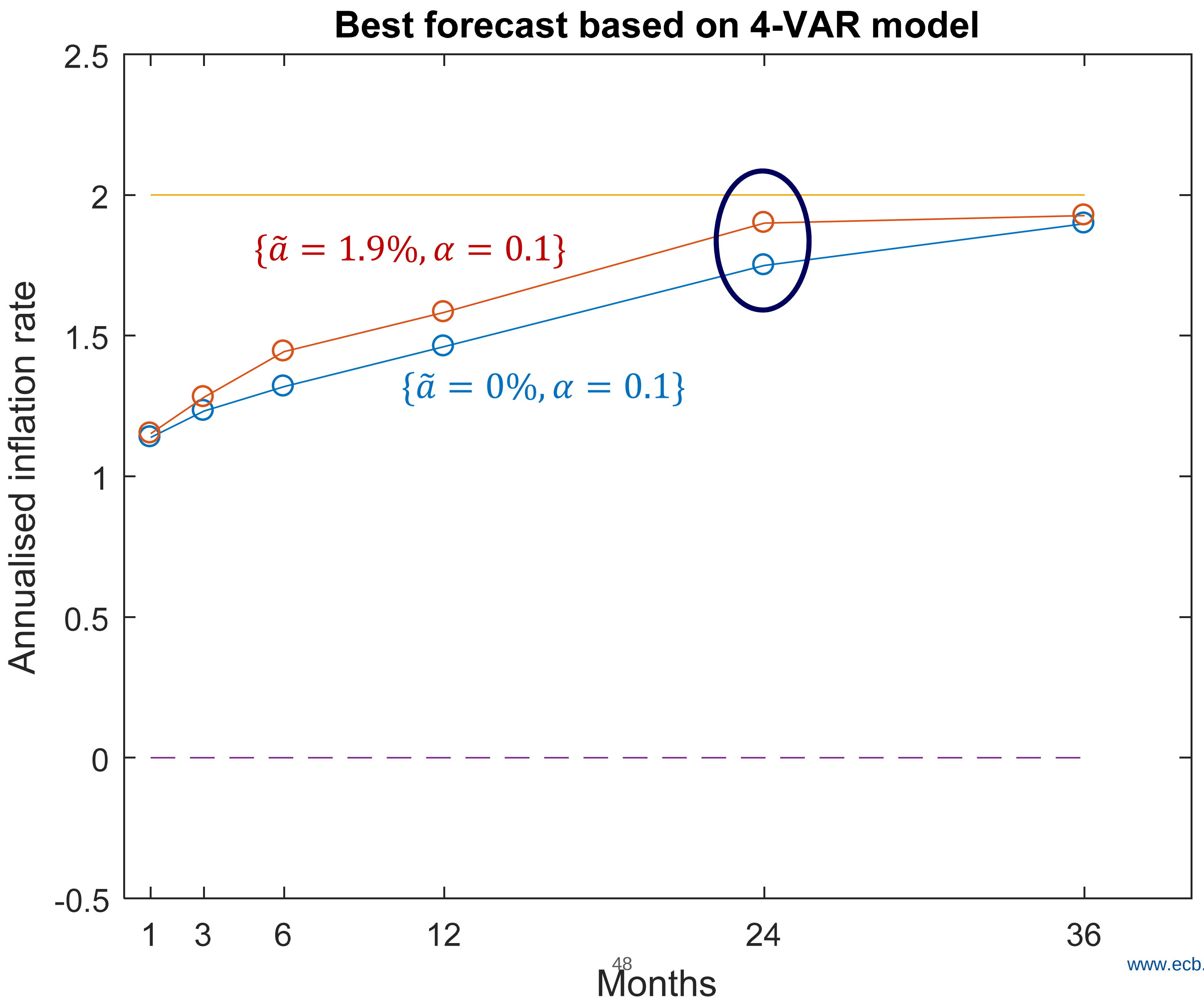
The frequentist statistical decision rule

Model selection under misspecification

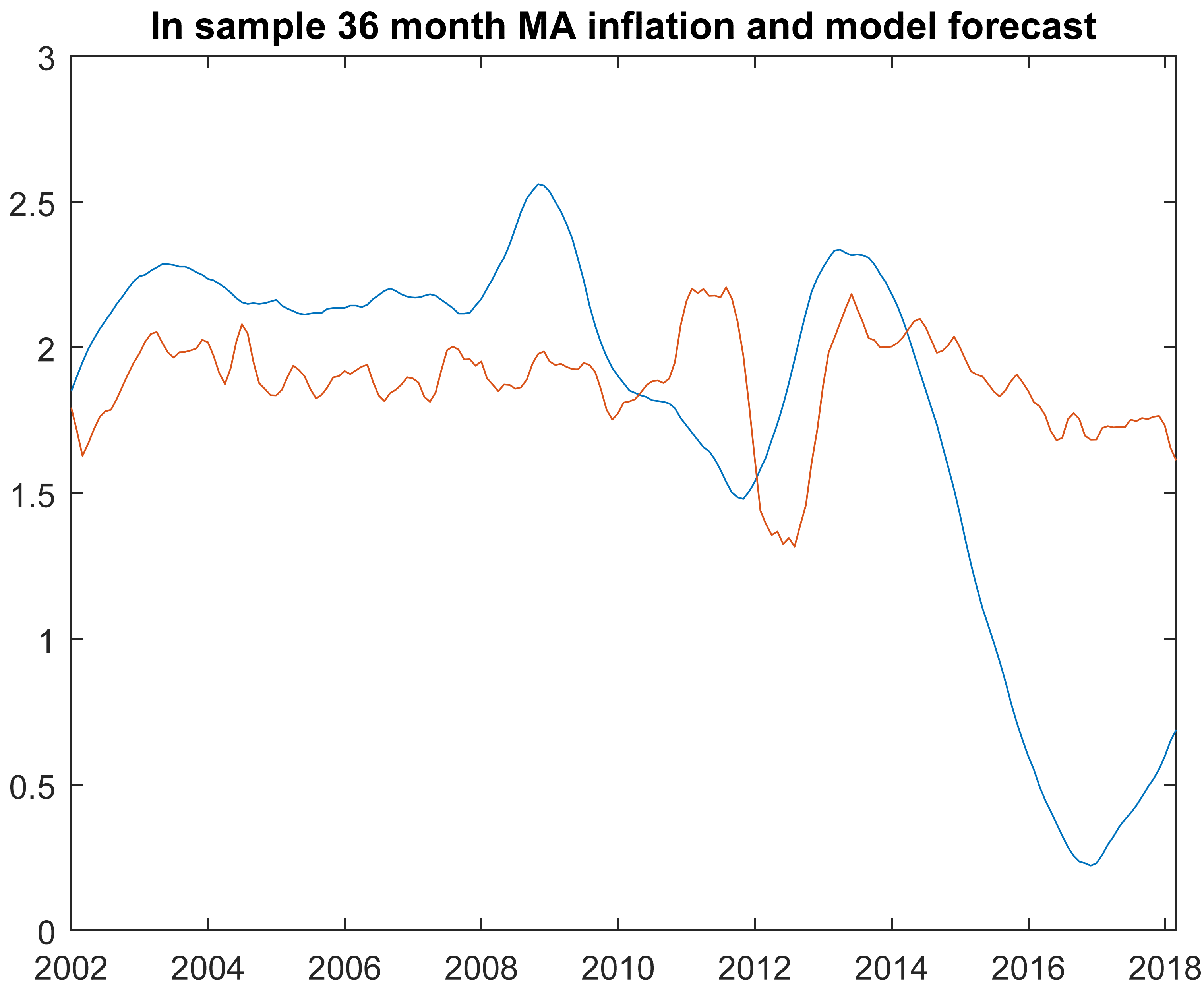
Empirical illustration

- Preliminaries
- **Results**

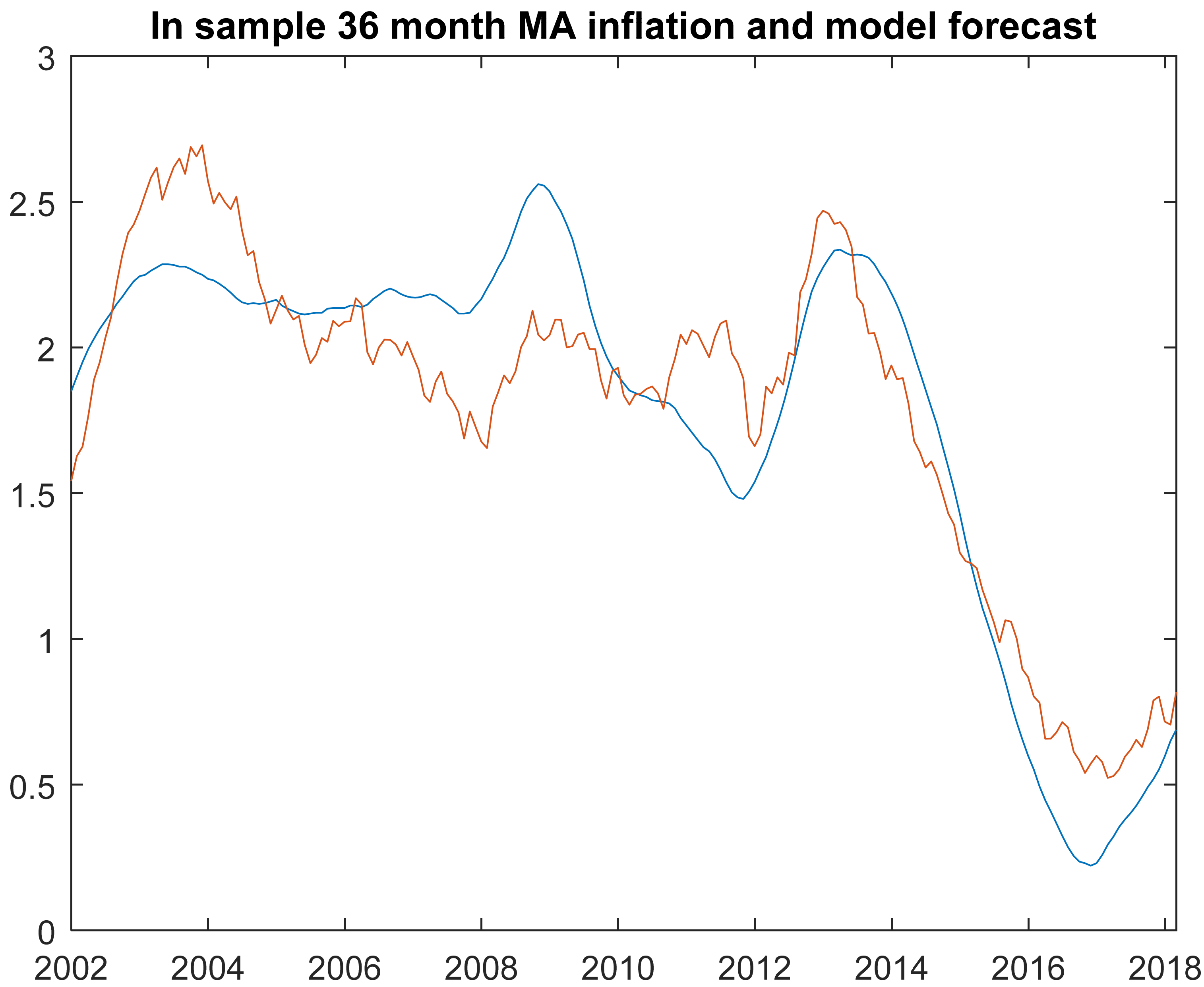
Euro area inflation forecast as of February 2018



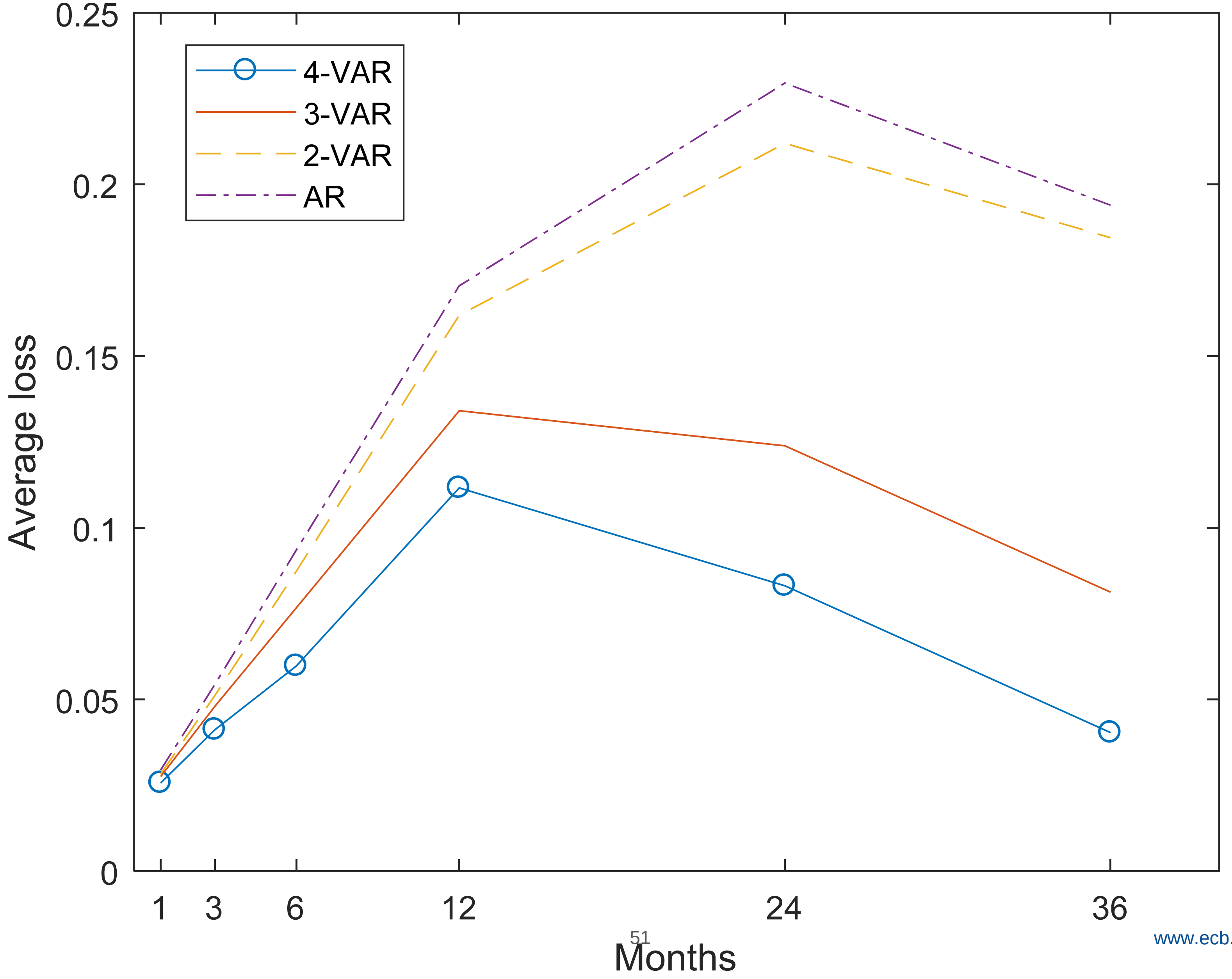
In sample performance of AR model



In sample performance of 4-VAR model



Average in sample loss



Conclusions

Conclusions

1. Separation b/w judgment and econometric modelling

Conclusions

1. Separation b/w judgment and econometric modelling

2. Methodology to incorporate judgment and select the least mis-specified model

Conclusions

1. Separation b/w judgment and econometric modelling

2. Methodology to incorporate judgment and select the least mis-specified model

3. Pay attention to unemployment, rather than core inflation