Simone Manganelli
European Central Bank

Selecting Models with Judgment

2nd “Forecasting at Central Banks” conference
Bank of England and King’s College London, 15 November, 2018

Disclaimer: Any views expressed are only the author’s own and do not necessarily reflect the views of the ECB or the Eurosystem
Question

Start from a judgmental forecast. How can it be statistically improved?
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) $L(\theta, \alpha) = 0.5(\theta - \alpha)^2$
3) Judgment: $\{\bar{\alpha}, \alpha\}$
4) $X = x$

FOC: $H_0: \nabla_\alpha L(\theta, \bar{\alpha}) = -\theta + \bar{\alpha} = 0$

Distribution of $\nabla_\alpha L(\hat{\theta}, \bar{\alpha}) = -X + \bar{\alpha}$ under $H_0: \theta = \bar{\alpha}$

$0 = -x + \hat{\alpha}^{ML}$  \( \xrightarrow{\text{under } H_0} \)  $-x + \hat{\alpha}$  \( \xrightarrow{\text{under } H_0} \)  $-x + \bar{\alpha}$

$1 - \alpha$ confidence interval
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: $\{\bar{a}, \alpha\}$
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: \{\bar{a}, \alpha\}
4) $X = x$

FOC: $H_0$: $\nabla_{\bar{a}} L(\theta, \bar{a}) = -\theta + \bar{a} = 0$
# The frequentist decision rule with judgment

<table>
<thead>
<tr>
<th>1) $X \sim N(\theta, 1)$</th>
<th>2) $L(\theta, a) = 0.5(\theta - a)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3) Judgment: ${\tilde{a}, \alpha}$</td>
<td>4) $X = x$</td>
</tr>
</tbody>
</table>

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

**Test statistic:** $-X + \tilde{a}$
The frequentist decision rule with judgment

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

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

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

\[ 0 = -x + \hat{a}^{ML} \]
The frequentist decision rule with judgment

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

FOC: \( H_0: \nabla_{\alpha} L(\theta, \tilde{\alpha}) = -\theta + \tilde{\alpha} = 0 \)
Test statistic: \( -X + \tilde{\alpha} \)

Distribution of \( \nabla_{\alpha} L(\hat{\theta}, \tilde{\alpha}) = -X + \tilde{\alpha} \) under \( H_0: \theta = \tilde{\alpha} \)

\[ 0 = -x + \hat{\alpha}^{ML} \]

\( 1 - \alpha \) confidence interval
The frequentist decision rule with judgment

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

FOC: \(H_0: \nabla_{\alpha} L(\theta, \tilde{\alpha}) = -\theta + \tilde{\alpha} = 0\)

Test statistic: \(-X + \tilde{\alpha}\)

Distribution of \(\nabla_{\alpha} L(\hat{\theta}, \tilde{\alpha}) = -X + \tilde{\alpha}\) under \(H_0: \theta = \tilde{\alpha}\)

\(0 = -x + \hat{\alpha}^{ML}\)

\(-x + \tilde{\alpha}\)

1 - \(\alpha\) confidence interval
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 $H_0: \theta = \tilde{a}$

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

$1 - \alpha$ confidence interval
The frequentist decision rule with judgment

\[ -x + \hat{\alpha} \quad 0 = -x + \hat{\alpha}^{ML} \quad -x + \hat{\alpha} \]

1 \( - \alpha \) confidence interval
The frequentist decision rule with judgment

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

1 − \( \alpha \) confidence interval
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:
\[
P_\theta[L(\theta, \delta(X)) > L(\theta, \tilde{a})] \leq \alpha
\]
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

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

\[
B^{*-\frac{1}{2}} A^* \sqrt{n} (\hat{\theta}(X^n) - \theta^*) \overset{A}{\rightarrow} N(0, I_p)
\]
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**

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

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

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:

$$\min_a L(\theta, a) = 0.5(x_{n,h}(\theta) - a)^2$$
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) $x_{n,h}(\hat{\theta}(X^n)) \sim N(x_{n,h}(\theta^*), \sigma^2)$

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

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

4) $X^n = x^n$

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

---

**Diagram:**

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

- **0 = $\nabla_a L(\hat{\theta}(x^n), \hat{\theta}^{ML})$**

- **$\nabla_a L(\hat{\theta}(x^n), \tilde{a})$**

- **1 $- \alpha$ confidence interval**
Model selection

Select the least misspecified model.
Model selection

Select the least misspecified model.

Minimise the Kullback-Leibler distance:

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

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

\[ \tilde{a} \equiv 24^{-1} E \left( \sum_{i=1}^{24} \pi_{n+i} \mid X^n \right) = 1.9\% \]
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

Inflation and core inflation in the euro area

www.ecb.europa.eu ©
Ingredients for forecasting

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

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

with zero restrictions:
\[ \rightarrow \text{ratio of parameters to data not too large} \]

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

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

<table>
<thead>
<tr>
<th></th>
<th>c</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>HICP</td>
<td>1</td>
<td>1010</td>
<td>0001</td>
<td>0000</td>
<td>0000</td>
<td>0000</td>
<td>1000</td>
</tr>
<tr>
<td>HICPX</td>
<td>0</td>
<td>1000</td>
<td>0000</td>
<td>0000</td>
<td>0000</td>
<td>0000</td>
<td>0100</td>
</tr>
<tr>
<td>UN</td>
<td>0</td>
<td>1000</td>
<td>0000</td>
<td>0000</td>
<td>0000</td>
<td>0000</td>
<td>0000</td>
</tr>
<tr>
<td>IP</td>
<td>1</td>
<td>0000</td>
<td>0001</td>
<td>0000</td>
<td>0000</td>
<td>0000</td>
<td>0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>0000</td>
<td>0100</td>
<td>0000</td>
<td>0000</td>
<td>0000</td>
<td>0011</td>
</tr>
<tr>
<td>0000</td>
<td>0000</td>
<td>1000</td>
<td>0010</td>
<td>0000</td>
<td>0000</td>
<td>0000</td>
</tr>
<tr>
<td>0000</td>
<td>0000</td>
<td>0010</td>
<td>0010</td>
<td>0000</td>
<td>0000</td>
<td>0000</td>
</tr>
<tr>
<td>0000</td>
<td>0001</td>
<td>0010</td>
<td>0010</td>
<td>0001</td>
<td>0000</td>
<td>0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>4</th>
<th>1</th>
<th>0</th>
<th>2</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>
The frequentist statistical decision rule
Model selection under misspecification
Empirical illustration
  – Preliminaries
  – Results
Euro area inflation forecast as of February 2018

Best forecast based on 4-VAR model

\{ \bar{a} = 1.9\%, \alpha = 0.1 \} \\
\{ \bar{a} = 0\%, \alpha = 0.1 \}
In sample performance of AR model

In sample 36 month MA inflation and model forecast
In sample performance of 4-VAR model
Average in sample loss

The graph shows the average loss over different months for different models:

- 4-VAR
- 3-VAR
- 2-VAR
- AR

The x-axis represents the number of months, and the y-axis represents the average loss. The graph illustrates the trend of average loss over time for each model.
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