Model-Acentric, Focused Bayesian Prediction

David Frazier, Ruben Loaiza Maya and Gael Martin

Monash University, Melbourne

BIRS Workshop, Mexico

November, 2018

Bayesian Prediction

Bayesian Prediction

- Distribution of interest is:

$$p\left(y_{T+1} \middle| \mathbf{y}_{1:T}\right)$$

Bayesian Prediction

- Distribution of interest is:

$$p(y_{T+1}|\mathbf{y}_{1:T})$$

Bayesian Prediction

- Distribution of interest is:

$$p(y_{T+1}|\mathbf{y}_{1:T}) = \int_{\boldsymbol{\theta}} p(y_{T+1}, \boldsymbol{\theta}|\mathbf{y}_{1:T}) d\boldsymbol{\theta}$$

Bayesian Prediction

- Distribution of interest is:

$$
\begin{aligned}
p(y_{T+1}|\mathbf{y}_{1:T}) &= \int_{\boldsymbol{\theta}} p(y_{T+1}, \boldsymbol{\theta}|\mathbf{y}_{1:T})d\boldsymbol{\theta} \\
&= \int_{\theta} p(y_{T+1}|\mathbf{y}_{1:T}, \boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y}_{1:T})d\boldsymbol{\theta}
\end{aligned}
$$

Bayesian Prediction

- Distribution of interest is:

$$
\begin{aligned}
p(y_{T+1}|\mathbf{y}_{1:T}) &= \int_{\boldsymbol{\theta}} p(y_{T+1}, \boldsymbol{\theta}|\mathbf{y}_{1:T}) d\boldsymbol{\theta} \\
&= \int_{\theta} p(y_{T+1}|\mathbf{y}_{1:T}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}_{1:T}) d\boldsymbol{\theta} \\
&= E_{\boldsymbol{\theta}|\mathbf{y}} \left[p(y_{T+1}|\mathbf{y}_{1:T}, \boldsymbol{\theta}) \right]
\end{aligned}
$$

Bayesian Prediction

- Distribution of interest is:

$$\begin{aligned} p(y_{T+1}|\mathbf{y}_{1:T}) &= \int_{\boldsymbol{\theta}} p(y_{T+1}, \boldsymbol{\theta}|\mathbf{y}_{1:T}) d\boldsymbol{\theta} \\ &= \int_{\theta} p(y_{T+1}|\mathbf{y}_{1:T}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}_{1:T}) d\boldsymbol{\theta} \\ &= E_{\boldsymbol{\theta}|\mathbf{y}} \left[p(y_{T+1}|\mathbf{y}_{1:T}, \boldsymbol{\theta}) \right] \end{aligned}$$

- **(Marginal)** predictive = expect. of **conditional** predictive

Bayesian Prediction

- Distribution of interest is:

$$
\begin{aligned}
p(y_{T+1}|\mathbf{y}_{1:T}) &= \int_{\boldsymbol{\theta}} p(y_{T+1}, \boldsymbol{\theta}|\mathbf{y}_{1:T}) d\boldsymbol{\theta} \\
&= \int_{\theta} p(y_{T+1}|\mathbf{y}_{1:T}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}_{1:T}) d\boldsymbol{\theta} \\
&= E_{\boldsymbol{\theta}|\mathbf{y}} \left[p(y_{T+1}|\mathbf{y}_{1:T}, \boldsymbol{\theta}) \right]
\end{aligned}
$$

- **(Marginal)** predictive = expect. of **conditional** predictive
- **Conditional** predictive reflects the **assumed DGP**

Bayesian Prediction

- Distribution of interest is:

$$
\begin{aligned}
p(y_{T+1}|\mathbf{y}_{1:T}) &= \int_{\boldsymbol{\theta}} p(y_{T+1}, \boldsymbol{\theta}|\mathbf{y}_{1:T}) d\boldsymbol{\theta} \\
&= \int_{\theta} p(y_{T+1}|\mathbf{y}_{1:T}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}_{1:T}) d\boldsymbol{\theta} \\
&= E_{\boldsymbol{\theta}|\mathbf{y}} \left[p(y_{T+1}|\mathbf{y}_{1:T}, \boldsymbol{\theta}) \right]
\end{aligned}
$$

- **(Marginal)** predictive = expect. of **conditional** predictive
- **Conditional** predictive reflects the **assumed DGP**
- As does $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$

Implementing Bayesian Prediction

Implementing Bayesian Prediction

- In the usual case where $E_{\boldsymbol{\theta}|\mathbf{y}}\left[p(y_{T+1}|\mathbf{y}_{1:T},\boldsymbol{\theta})\right]$ cannot be evaluated **analytically**

Implementing Bayesian Prediction

- In the usual case where $E_{\boldsymbol{\theta}|\mathbf{y}}\left[p(y_{T+1}|\mathbf{y}_{1:T}, \boldsymbol{\theta})\right]$ cannot be evaluated **analytically**
- Take M draws from $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ (via a Markov chain Monte Carlo algorithm, say)

Implementing Bayesian Prediction

- In the usual case where $E_{\boldsymbol{\theta}|\mathbf{y}}\left[p(y_{T+1}|\mathbf{y}_{1:T},\boldsymbol{\theta})\right]$ cannot be evaluated **analytically**
- Take M draws from $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ (via a Markov chain Monte Carlo algorithm, say)

Implementing Bayesian Prediction

- In the usual case where $E_{\boldsymbol{\theta}|\mathbf{y}}\left[p(y_{T+1}|\mathbf{y}_{1:T},\boldsymbol{\theta})\right]$ cannot be evaluated **analytically**
- Take M draws from $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ (via a Markov chain Monte Carlo algorithm, say)
- And **estimate** $p(y_{T+1}|\mathbf{y}_{1:T})$ as

Implementing Bayesian Prediction

- In the usual case where $E_{\boldsymbol{\theta}|\mathbf{y}}\left[p(y_{T+1}|\mathbf{y}_{1:T},\boldsymbol{\theta})\right]$ cannot be evaluated **analytically**
- Take M draws from $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ (via a Markov chain Monte Carlo algorithm, say)
- And **estimate** $p(y_{T+1}|\mathbf{y}_{1:T})$ as
  1. either:

$$\widehat{p}(y_{T+1}|\mathbf{y}_{1:T}) = \frac{1}{M}\sum_{i-1}^{M} p(y_{T+1}|\mathbf{y}_{1:T},\boldsymbol{\theta}^{(i)})$$

Implementing Bayesian Prediction

- In the usual case where $E_{\boldsymbol{\theta}|\mathbf{y}}\left[p(y_{T+1}|\mathbf{y}_{1:T},\boldsymbol{\theta})\right]$ cannot be evaluated **analytically**
- Take M draws from $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ (via a Markov chain Monte Carlo algorithm, say)
- And **estimate** $p(y_{T+1}|\mathbf{y}_{1:T})$ as
  1. either:
  $$\widehat{p}(y_{T+1}|\mathbf{y}_{1:T}) = \frac{1}{M}\sum_{i-1}^{M} p(y_{T+1}|\mathbf{y}_{1:T},\boldsymbol{\theta}^{(i)})$$
  2. or: $\widehat{p}(y_{T+1}|\mathbf{y}_{1:T})$ constructed from draws of $y_{T+1}^{(i)}$ simulated from $p(y_{T+1}|\mathbf{y}_{1:T},\boldsymbol{\theta}^{(i)})$

Implementing Bayesian Prediction

- In the usual case where $E_{\boldsymbol{\theta}|\mathbf{y}}\left[p(y_{T+1}|\mathbf{y}_{1:T},\boldsymbol{\theta})\right]$ cannot be evaluated **analytically**
- Take M draws from $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ (via a Markov chain Monte Carlo algorithm, say)
- And **estimate** $p(y_{T+1}|\mathbf{y}_{1:T})$ as
  1. either:
  $$\widehat{p}(y_{T+1}|\mathbf{y}_{1:T}) = \frac{1}{M}\sum_{i-1}^{M} p(y_{T+1}|\mathbf{y}_{1:T},\boldsymbol{\theta}^{(i)})$$
  2. or: $\widehat{p}(y_{T+1}|\mathbf{y}_{1:T})$ constructed from draws of $y_{T+1}^{(i)}$ simulated from $p(y_{T+1}|\mathbf{y}_{1:T},\boldsymbol{\theta}^{(i)})$
- i.e. MCMC $\Rightarrow$ **exact Bayesian prediction**

Implementing Bayesian Prediction

- In the usual case where $E_{\boldsymbol{\theta}|\mathbf{y}}\left[p(y_{T+1}|\mathbf{y}_{1:T},\boldsymbol{\theta})\right]$ cannot be evaluated **analytically**
- Take M draws from $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ (via a Markov chain Monte Carlo algorithm, say)
- And **estimate** $p(y_{T+1}|\mathbf{y}_{1:T})$ as
  1. either:
  $$\widehat{p}(y_{T+1}|\mathbf{y}_{1:T}) = \frac{1}{M}\sum_{i-1}^{M} p(y_{T+1}|\mathbf{y}_{1:T},\boldsymbol{\theta}^{(i)})$$
  2. or: $\widehat{p}(y_{T+1}|\mathbf{y}_{1:T})$ constructed from draws of $y_{T+1}^{(i)}$ simulated from $p(y_{T+1}|\mathbf{y}_{1:T},\boldsymbol{\theta}^{(i)})$
- i.e. MCMC $\Rightarrow$ **exact Bayesian prediction**
  - (up to simulation error)

Achilles Heels!

Achilles Heels!

1. What happens when we can't generate an MCMC chain because $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ is inaccessible?

Achilles Heels!

1. What happens when we can't generate an MCMC chain because $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ is inaccessible?
   - $\Rightarrow$ **exact** Bayesian prediction not feasible

Achilles Heels!

1. What happens when we can't generate an MCMC chain because $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ is inaccessible?
   - $\Rightarrow$ **exact** Bayesian prediction not feasible
   - $\Rightarrow$ **Frazier et al. (2018)**: "*Approximate Bayesian Forecasting*"

Achilles Heels!

1. What happens when we can't generate an MCMC chain because $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ is inaccessible?
   - $\Rightarrow$ **exact** Bayesian prediction not feasible
   - $\Rightarrow$ **Frazier et al. (2018)**: "*Approximate Bayesian Forecasting*"
2. What happens when we acknowledge that the **DGP** used to construct $p(y_{T+1}|\mathbf{y}_{1:T})$ **misspecified**?

Achilles Heels!

1. What happens when we can't generate an MCMC chain because $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ is inaccessible?
   - $\Rightarrow$ **exact** Bayesian prediction not feasible
   - $\Rightarrow$ **Frazier et al. (2018)**: "*Approximate Bayesian Forecasting*"
2. What happens when we acknowledge that the **DGP** used to construct $p(y_{T+1}|\mathbf{y}_{1:T})$ **misspecified**?
   - This impinges on $p(y_{T+1}|\mathbf{y}_{1:T})$ via its two components:

$$p(y_{T+1}|\mathbf{y}_{1:T}) = \int_{\theta} p(y_{T+1}|\boldsymbol{\theta}, \mathbf{y}_{1:T})p(\boldsymbol{\theta}|\mathbf{y}_{1:T})d\boldsymbol{\theta} \text{ and}$$

Achilles Heels!

1. What happens when we can't generate an MCMC chain because $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ is inaccessible?
   - $\Rightarrow$ **exact** Bayesian prediction not feasible
   - $\Rightarrow$ **Frazier et al. (2018)**: "*Approximate Bayesian Forecasting*"
2. What happens when we acknowledge that the **DGP** used to construct $p(y_{T+1}|\mathbf{y}_{1:T})$ **misspecified**?
   - This impinges on $p(y_{T+1}|\mathbf{y}_{1:T})$ via its two components:

$$p(y_{T+1}|\mathbf{y}_{1:T}) = \int_{\theta} p(y_{T+1}|\boldsymbol{\theta}, \mathbf{y}_{1:T})p(\boldsymbol{\theta}|\mathbf{y}_{1:T})d\boldsymbol{\theta} \text{ and}$$

Achilles Heels!

1. What happens when we can't generate an MCMC chain because $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ is inaccessible?
   - $\Rightarrow$ **exact** Bayesian prediction not feasible
   - $\Rightarrow$ **Frazier et al. (2018)**: "*Approximate Bayesian Forecasting*"
2. What happens when we acknowledge that the **DGP** used to construct $p(y_{T+1}|\mathbf{y}_{1:T})$ **misspecified**?
   - This impinges on $p(y_{T+1}|\mathbf{y}_{1:T})$ via its two components:

$$p(y_{T+1}|\mathbf{y}_{1:T}) = \int_{\theta} p(y_{T+1}|\boldsymbol{\theta}, \mathbf{y}_{1:T})p(\boldsymbol{\theta}|\mathbf{y}_{1:T})d\boldsymbol{\theta} \text{ and}$$

   - The **conditional** predictive: $p(y_{T+1}|\boldsymbol{\theta}, \mathbf{y}_{1:T})$

Achilles Heels!

1. What happens when we can't generate an MCMC chain because $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ is inaccessible?
   - $\Rightarrow$ **exact** Bayesian prediction not feasible
   - $\Rightarrow$ **Frazier et al. (2018)**: "*Approximate Bayesian Forecasting*"
2. What happens when we acknowledge that the **DGP** used to construct $p(y_{T+1}|\mathbf{y}_{1:T})$ **misspecified**?
   - This impinges on $p(y_{T+1}|\mathbf{y}_{1:T})$ via its two components:

$$p(y_{T+1}|\mathbf{y}_{1:T}) = \int_{\theta} p(y_{T+1}|\boldsymbol{\theta}, \mathbf{y}_{1:T}) p(\boldsymbol{\theta}|\mathbf{y}_{1:T}) d\boldsymbol{\theta} \text{ and}$$

   - The **conditional** predictive: $p(y_{T+1}|\boldsymbol{\theta}, \mathbf{y}_{1:T})$
   - and $p(\boldsymbol{\theta}|\mathbf{y}_{1:T}) \propto p(\mathbf{y}_{1:T}|\boldsymbol{\theta}) \times p(\boldsymbol{\theta})$

Achilles Heels!

1. What happens when we can't generate an MCMC chain because $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ is inaccessible?
   - $\Rightarrow$ **exact** Bayesian prediction not feasible
   - $\Rightarrow$ **Frazier et al. (2018)**: "*Approximate Bayesian Forecasting*"
2. What happens when we acknowledge that the **DGP** used to construct $p(y_{T+1}|\mathbf{y}_{1:T})$ **misspecified**?
   - This impinges on $p(y_{T+1}|\mathbf{y}_{1:T})$ via its two components:
   $$p(y_{T+1}|\mathbf{y}_{1:T}) = \int_{\theta} p(y_{T+1}|\boldsymbol{\theta}, \mathbf{y}_{1:T})p(\boldsymbol{\theta}|\mathbf{y}_{1:T})d\boldsymbol{\theta} \text{ and}$$
   - The **conditional** predictive: $p(y_{T+1}|\boldsymbol{\theta}, \mathbf{y}_{1:T})$
   - and $p(\boldsymbol{\theta}|\mathbf{y}_{1:T}) \propto p(\mathbf{y}_{1:T}|\boldsymbol{\theta}) \times p(\boldsymbol{\theta})$
   - In what sense does $p(y_{T+1}|\mathbf{y}_{1:T})$ remain the gold standard?

A New Paradigm for Bayesian Prediction

A New Paradigm for Bayesian Prediction

- Appropriate for the realistic setting in which the **true DGP is unknown**

A New Paradigm for Bayesian Prediction

- Appropriate for the realistic setting in which the **true DGP is unknown**
- Define $\mathcal{P}$ as the class of **conditional predictives** that we believe **could** have generated the data

A New Paradigm for Bayesian Prediction

- Appropriate for the realistic setting in which the **true DGP is unknown**
- Define $\mathcal{P}$ as the class of **conditional predictives** that we believe **could** have generated the data
- With elements:

$$P \in \mathcal{P}$$

A New Paradigm for Bayesian Prediction

- Appropriate for the realistic setting in which the **true DGP is unknown**
- Define $\mathcal{P}$ as the class of **conditional predictives** that we believe **could** have generated the data
- With elements:

$$P \in \mathcal{P}$$

- where P denotes some conditional distribution

A New Paradigm for Bayesian Prediction

- Appropriate for the realistic setting in which the **true DGP is unknown**
- Define $\mathcal{P}$ as the class of **conditional predictives** that we believe **could** have generated the data
- With elements:

$$P \in \mathcal{P}$$

- where P denotes some conditional distribution
- In principle, $\mathcal{P}$ may be a class of:

A New Paradigm for Bayesian Prediction

- Appropriate for the realistic setting in which the **true DGP is unknown**
- Define $\mathcal{P}$ as the class of **conditional predictives** that we believe **could** have generated the data
- With elements:

$$P \in \mathcal{P}$$

- where P denotes some conditional distribution
- In principle, $\mathcal{P}$ may be a class of:
    - distributions, $P_{\boldsymbol{\theta}}$ say, associated with a **given parametric** model

A New Paradigm for Bayesian Prediction

- Appropriate for the realistic setting in which the **true DGP is unknown**
- Define $\mathcal{P}$ as the class of **conditional predictives** that we believe **could** have generated the data
- With elements:
$$P \in \mathcal{P}$$
- where P denotes some conditional distribution
- In principle, $\mathcal{P}$ may be a class of:
  - distributions, $P_{\boldsymbol{\theta}}$ say, associated with a **given parametric** model
  - weighted combinations of predictives associated with **different parametric** models

A New Paradigm for Bayesian Prediction

- Appropriate for the realistic setting in which the **true DGP is unknown**
- Define $\mathcal{P}$ as the class of **conditional predictives** that we believe **could** have generated the data
- With elements:

$$P \in \mathcal{P}$$

- where P denotes some conditional distribution
- In principle, $\mathcal{P}$ may be a class of:
  - distributions, $P_{\boldsymbol{\theta}}$ say, associated with a **given parametric** model
  - weighted combinations of predictives associated with **different parametric** models
  - **non-parametric** conditional distributions

A New Paradigm for Bayesian Prediction

- Define predictive **a proper scoring rule:** $S(P, y_{t+1})$

A New Paradigm for Bayesian Prediction

- Define predictive **a proper scoring rule:** $S(P, y_{t+1})$
- with expectation, under the **truth,** F, as:

$$\mathcal{S}(P, F) = \mathbb{E}_F \left[S(P, Y_{t+1}) \right]$$

A New Paradigm for Bayesian Prediction

- Define predictive **a proper scoring rule:** $S(P, y_{t+1})$
- with expectation, under the **truth,** F, as:

$$\mathcal{S}(P, F) = \mathbb{E}_F \left[S(P, Y_{t+1}) \right]$$

- The map $P \mapsto -\mathcal{S}(P, F)$ defines a **loss function** over the models in $\mathcal{P}$

A New Paradigm for Bayesian Prediction

- Define predictive **a proper scoring rule:** $S(P, y_{t+1})$
- with expectation, under the **truth,** F, as:

$$\mathcal{S}(P, F) = \mathbb{E}_F \left[S(P, Y_{t+1}) \right]$$

- The map $P \mapsto -\mathcal{S}(P, F)$ defines a **loss function** over the models in $\mathcal{P}$
- Aim is to **focus** on the elements of $\mathcal{P}$ that **minimize this loss**

Model-Acentric, Focused Bayesian Prediction

Model-Acentric, Focused Bayesian Prediction

- $\Rightarrow$ puts **focus on** (user-defined) minimizing out-of-sample loss

Model-Acentric, Focused Bayesian Prediction

- $\Rightarrow$ puts **focus on** (user-defined) minimizing out-of-sample loss
- $\Rightarrow$ takes **focus away** from a particular (wrong!) **model**

Model-Acentric, Focused Bayesian Prediction

- $\Rightarrow$ puts **focus on** (user-defined) minimizing out-of-sample loss
- $\Rightarrow$ takes **focus away** from a particular (wrong!) **model**
- Partition the sample: $y_1, y_2, ..., y_T$ into:

Model-Acentric, Focused Bayesian Prediction

- $\Rightarrow$ puts **focus on** (user-defined) minimizing out-of-sample loss
- $\Rightarrow$ takes **focus away** from a particular (wrong!) **model**
- Partition the sample: $y_1, y_2, ..., y_T$ into:
  - A **training** set: $\mathcal{D} = \{y_t;\, 1 \leq t \leq \tau\}$

Model-Acentric, Focused Bayesian Prediction

- $\Rightarrow$ puts **focus on** (user-defined) minimizing out-of-sample loss
- $\Rightarrow$ takes **focus away** from a particular (wrong!) **model**
- Partition the sample: $y_1, y_2, ..., y_T$ into:
  - A **training** set: $\mathcal{D} = \{y_t;\, 1 \leq t \leq \tau\}$
  - A **test** set: $\mathcal{T} = \{y_t;\, \tau + 1 \leq t \leq \tau + n = T\}$

Model-Acentric, Focused Bayesian Prediction

- $\Rightarrow$ puts **focus on** (user-defined) minimizing out-of-sample loss
- $\Rightarrow$ takes **focus away** from a particular (wrong!) **model**
- Partition the sample: $y_1, y_2, ..., y_T$ into:
  - A **training** set: $\mathcal{D} = \{y_t;\ 1 \leq t \leq \tau\}$
  - A **test** set: $\mathcal{T} = \{y_t;\ \tau + 1 \leq t \leq \tau + n = T\}$
- **Fit** P on $\mathcal{D} \Rightarrow \widehat{P}_{1:t} = \widehat{P}[.|y_{1:t}]$

Model-Acentric, Focused Bayesian Prediction

- $\Rightarrow$ puts **focus on** (user-defined) minimizing out-of-sample loss
- $\Rightarrow$ takes **focus away** from a particular (wrong!) **model**
- Partition the sample: $y_1, y_2, ..., y_T$ into:
  - A **training** set: $\mathcal{D} = \{y_t;\ 1 \leq t \leq \tau\}$
  - A **test** set: $\mathcal{T} = \{y_t;\ \tau + 1 \leq t \leq \tau + n = T\}$
- **Fit** P on $\mathcal{D} \Rightarrow \widehat{P}_{1:t} = \widehat{P}[.|y_{1:t}]$
- Use $\mathcal{T}$ (and **expanding** $\mathcal{D}$) to **compute**:

$$S_n(P, F) = \frac{1}{n} \sum_{i=0}^{n-1} S(\widehat{P}_{1:(\tau+i)}, y_{(\tau+i)+1})$$

Model-Acentric, Focused Bayesian Prediction

- $\Rightarrow$ puts **focus on** (user-defined) minimizing out-of-sample loss
- $\Rightarrow$ takes **focus away** from a particular (wrong!) **model**
- Partition the sample: $y_1, y_2, ..., y_T$ into:
  - A **training** set: $\mathcal{D} = \{y_t;\ 1 \leq t \leq \tau\}$
  - A **test** set: $\mathcal{T} = \{y_t;\ \tau + 1 \leq t \leq \tau + n = T\}$
- **Fit** P on $\mathcal{D} \Rightarrow \widehat{P}_{1:t} = \widehat{P}[.|y_{1:t}]$
- Use $\mathcal{T}$ (and **expanding** $\mathcal{D}$) to **compute**:

$$S_n(P, F) = \frac{1}{n} \sum_{i=0}^{n-1} S(\widehat{P}_{1:(\tau+i)}, y_{(\tau+i)+1})$$

- as an estimate of $\mathcal{S}(P, F) = \mathbb{E}_F \left[S(P, Y_{t+1})\right]$

Focused Bayesian Prediction (FBP)

- Now define a prior over the elements of $\mathcal{P}$: $\Pi(P)$

Focused Bayesian Prediction (FBP)

- Now define a prior over the elements of $\mathcal{P} : \Pi(P)$
- Update **prior** $\Pi(P)$ to **posterior** $\Pi(P|.)$ according to **predictive performance** over the test set, $\mathcal{T}$

Focused Bayesian Prediction (FBP)

- Now define a prior over the elements of $\mathcal{P} : \Pi(P)$
- Update **prior** $\Pi(P)$ to **posterior** $\Pi(P|.)$ according to **predictive performance** over the test set, $\mathcal{T}$
- i.e. $\Pi(P|.)$ is tuned, or **calibrated,** to assign high posterior mass to elements of $\mathcal{P}$ with **high predictive accuracy**

Focused Bayesian Prediction (FBP)

- Now define a prior over the elements of $\mathcal{P} : \Pi(P)$
- Update **prior** $\Pi(P)$ to **posterior** $\Pi(P|.)$ according to **predictive performance** over the test set, $\mathcal{T}$
- i.e. $\Pi(P|.)$ is tuned, or **calibrated,** to assign high posterior mass to elements of $\mathcal{P}$ with **high predictive accuracy**
- $\Leftrightarrow$ **small loss**

Focused Bayesian Prediction (FBP)

- Now define a prior over the elements of $\mathcal{P} : \Pi(P)$
- Update **prior** $\Pi(P)$ to **posterior** $\Pi(P|.)$ according to **predictive performance** over the test set, $\mathcal{T}$
- i.e. $\Pi(P|.)$ is tuned, or **calibrated,** to assign high posterior mass to elements of $\mathcal{P}$ with **high predictive accuracy**
- $\Leftrightarrow$ **small loss**
- $\Rightarrow \Pi(P|.)$ is '**focused' on** elements of $\mathcal{P}$ that **minimize this particular loss**

Focused Bayesian Prediction (FBP)

- **FBP Algorithm:**

Focused Bayesian Prediction (FBP)

- **FBP Algorithm:**

Focused Bayesian Prediction (FBP)

- **FBP Algorithm:**

$$1. \quad \text{Draw } P^i \text{ from } \Pi(P), \qquad i = 1, 2, ..., N$$

Focused Bayesian Prediction (FBP)

- **FBP Algorithm:**

1. Draw P^i from $\Pi(P)$, $\quad i = 1, 2, ..., N$
2. Compute $\widehat{P}^i_{1:t}$ $\quad$ using $\mathcal{D}$ and P^i

Focused Bayesian Prediction (FBP)

- **FBP Algorithm:**

1.	Draw P^i from $\Pi(P)$,	$i = 1, 2, ..., N$
2.	Compute $\widehat{P}^i_{1:t}$	using $\mathcal{D}$ and P^i
3.	Generate $s = S_n(\widehat{P}^i_{1:t}, F)$	over test set $\mathcal{T}$

Focused Bayesian Prediction (FBP)

- **FBP Algorithm:**

1.	Draw P^i from $\Pi(P)$,	$i = 1, 2, ..., N$
2.	Compute $\widehat{P}^i_{1:t}$	using $\mathcal{D}$ and P^i
3.	Generate $s = S_n(\widehat{P}^i_{1:t}, F)$	over test set $\mathcal{T}$
3.	For each $i = 1, 2, ..., N$	accept $\widehat{P}^i_{1:t}$ if $s \geq \varepsilon_n$

Focused Bayesian Prediction (FBP)

- **FBP Algorithm:**

1.	Draw P^i from $\Pi(P)$,	$i = 1, 2, ..., N$
2.	Compute $\widehat{P}^i_{1:t}$	using $\mathcal{D}$ and P^i
3.	Generate $s = S_n(\widehat{P}^i_{1:t}, F)$	over test set $\mathcal{T}$
3.	For each $i = 1, 2, ..., N$	accept $\widehat{P}^i_{1:t}$ if $s \geq \varepsilon_n$

- **Conditional** on $y_{1:t}$, and the **observed** $s = S_n(\widehat{P}^i_{1:t}, F)$

Focused Bayesian Prediction

- This **likelihood-free algorithm** produces *i.i.d.* draws from the **posterior** distribution with pdf:

Focused Bayesian Prediction

- This **likelihood-free algorithm** produces $i.i.\dot{d}.$ draws from the **posterior** distribution with pdf:

Focused Bayesian Prediction

- This **likelihood-free algorithm** produces $i.i.d.$ draws from the **posterior** distribution with pdf:

$$\pi_{\varepsilon_n}[P|s] = \frac{\pi(P)g_n[s|P]\mathbb{I}\{s \in A_{\varepsilon_n}\}}{\int_{\mathcal{P}} \pi(P)g_n[s|P]\mathbb{I}\{s \in A_{\varepsilon_n}\}dP}$$

$$A_{\varepsilon_n} = \{P \in \mathcal{P},\ s \in \mathcal{B} : s \sim G_n(.|P),\ \text{and } s \geq \varepsilon_n\}$$

Focused Bayesian Prediction

- This **likelihood-free algorithm** produces $i.i.\dot{d}.$ draws from the **posterior** distribution with pdf:

$$\pi_{\varepsilon_n}[P|s] = \frac{\pi(P)g_n[s|P]\mathbb{I}\{s \in A_{\varepsilon_n}\}}{\int\limits_{\mathcal{P}} \pi(P)g_n[s|P]\mathbb{I}\{s \in A_{\varepsilon_n}\}dP}$$

$$A_{\varepsilon_n} = \{P \in \mathcal{P},\ s \in \mathcal{B} : s \sim G_n(.|P),\ \text{and } s \geq \varepsilon_n\}$$

- where $G_n(.|P)$ is the **distribution** of $S_n(P, F)$, under F, with **pdf** $g_n[s|P]$

Focused Bayesian Prediction

- Draws produce a **nonparametric** estimate of $\pi_{\varepsilon_n}[P|s]$ that:

Focused Bayesian Prediction

- Draws produce a **nonparametric** estimate of $\pi_{\varepsilon_n}[P|s]$ that:
  - Does **not** require a closed-form for $g_n[s|P]$ (for $F \in \mathcal{P}$)

Focused Bayesian Prediction

- Draws produce a **nonparametric** estimate of $\pi_{\varepsilon_n}[P|s]$ that:
  - Does **not** require a closed-form for $g_n[s|P]$ (for $F \in \mathcal{P}$)
  - Does **not** require assumption that $F \in \mathcal{P}$

Focused Bayesian Prediction

- Draws produce a **nonparametric** estimate of $\pi_{\varepsilon_n}[P|s]$ that:
  - Does **not** require a closed-form for $g_n[s|P]$ (for $F \in \mathcal{P}$)
  - Does **not** require assumption that $F \in \mathcal{P}$
  - i.e. explicitly accommodates **model mis-specification**

Focused Bayesian Prediction

- Draws produce a **nonparametric** estimate of $\pi_{\varepsilon_n}[P|s]$ that:
  - Does **not** require a closed-form for $g_n[s|P]$ (for $F \in \mathcal{P}$)
  - Does **not** require assumption that $F \in \mathcal{P}$
  - i.e. explicitly accommodates **model mis-specification**
- Different (problem-specific) measures of **loss** $\Rightarrow$ different $\pi_{\varepsilon_n}[P|s]$

Focused Bayesian Prediction

- Draws produce a **nonparametric** estimate of $\pi_{\varepsilon_n}[P|s]$ that:
  - Does **not** require a closed-form for $g_n[s|P]$ (for $F \in \mathcal{P}$)
  - Does **not** require assumption that $F \in \mathcal{P}$
  - i.e. explicitly accommodates **model mis-specification**
- Different (problem-specific) measures of **loss** $\Rightarrow$ different $\pi_{\varepsilon_n}[P|s]$
- Different choices for ε_n

Focused Bayesian Prediction

- Draws produce a **nonparametric** estimate of $\pi_{\varepsilon_n}[P|s]$ that:
  - Does **not** require a closed-form for $g_n[s|P]$ (for $F \in \mathcal{P}$)
  - Does **not** require assumption that $F \in \mathcal{P}$
  - i.e. explicitly accommodates **model mis-specification**
- Different (problem-specific) measures of **loss** $\Rightarrow$ different $\pi_{\varepsilon_n}[P|s]$
- Different choices for ε_n
- $\Rightarrow$ different **aversion to (or tolerance of) loss**

Preliminary Theoretical Results

Preliminary Theoretical Results

- **Theorem 1: Posterior Concentration** (of $\Pi_{\varepsilon_n}[P|s]$) **:**

Preliminary Theoretical Results

- **Theorem 1: Posterior Concentration** (of $\Pi_{\varepsilon_n}[P|s]$) **:**
- Define:

$$P^* = \arg\max_{P \in \mathcal{P}} \mathcal{S}(P, F) \text{ with } \varepsilon^* = \mathcal{S}(P^*, F)$$

Preliminary Theoretical Results

- **Theorem 1: Posterior Concentration** (of $\Pi_{\varepsilon_n}[P|s]$) **:**
- Define:

$$P^* = \arg\max_{P \in \mathcal{P}} \mathcal{S}(P, F) \text{ with } \varepsilon^* = \mathcal{S}(P^*, F)$$

- As $\varepsilon_n \to \varepsilon^*$ (and under other mild conditions.....):

Preliminary Theoretical Results

- **Theorem 1: Posterior Concentration** (of $\Pi_{\varepsilon_n}[P|s]$) **:**
- Define:

$$P^* = \arg\max_{P \in \mathcal{P}} \mathcal{S}(P, F) \text{ with } \varepsilon^* = \mathcal{S}(P^*, F)$$

- As $\varepsilon_n \rightarrow \varepsilon^*$ (and under other mild conditions.....):

Preliminary Theoretical Results

- **Theorem 1: Posterior Concentration** (of $\Pi_{\varepsilon_n}[P|s]$) **:**
- Define:

$$P^* = \arg\max_{P \in \mathcal{P}} \mathcal{S}(P, F) \text{ with } \varepsilon^* = \mathcal{S}(P^*, F)$$

- As $\varepsilon_n \rightarrow \varepsilon^*$ (and under other mild conditions.....):

$$\Pi_{\varepsilon_n}[|\mathcal{S}(P, F) - \mathcal{S}(P^*, F)| > \delta_n | s] \underset{n \rightarrow \infty}{\rightarrow} 0$$

Preliminary Theoretical Results

- **Theorem 1: Posterior Concentration** (of $\Pi_{\varepsilon_n}[P|s]$) **:**
- Define:

$$P^* = \arg\max_{P\in\mathcal{P}} \mathcal{S}(P, F) \text{ with } \varepsilon^* = \mathcal{S}(P^*, F)$$

- As $\varepsilon_n \to \varepsilon^*$ (and under other mild conditions.....):

$$\Pi_{\varepsilon_n}[|\mathcal{S}(P, F) - \mathcal{S}(P^*, F)| > \delta_n|s] \underset{n\to\infty}{\to} 0$$

- $\Rightarrow$ **posterior distribution** of the expected score of $P \in \mathcal{P}$ **concentrates onto** the **maximum** expected score possible under F

Preliminary Theoretical Results

- In the case where $\mathcal{P}$ defines a class of **parametric** predictives (and under added assumptions more generally)

Preliminary Theoretical Results

- In the case where $\mathcal{P}$ defines a class of **parametric** predictives (and under added assumptions more generally)
- $\Rightarrow$

$$\Pi_{\varepsilon_n}[|P - P^*| \;>\; \delta_n | s] \underset{n\to\infty}{\longrightarrow} 0$$

Preliminary Theoretical Results

- In the case where $\mathcal{P}$ defines a class of **parametric** predictives (and under added assumptions more generally)
- $\Rightarrow$

$$\Pi_{\varepsilon_n}[|P - P^*| \; > \; \delta_n|s] \underset{n\to\infty}{\longrightarrow} 0$$

- $\Rightarrow$ **posterior distribution** of $P \in \mathcal{P}$ **concentrates onto** the predictive distribution that:

Preliminary Theoretical Results

- In the case where $\mathcal{P}$ defines a class of **parametric** predictives (and under added assumptions more generally)
- $\Rightarrow$

$$\Pi_{\varepsilon_n}\left[\left|P - P^*\right| \ > \ \delta_n | s\right] \underset{n\rightarrow\infty}{\rightarrow} 0$$

- $\Rightarrow$ **posterior distribution** of $P \in \mathcal{P}$ **concentrates onto** the predictive distribution that:
- **maximizes the expected score** $\Leftrightarrow$ **minimizes loss**

Preliminary Theoretical Results

- So $\Pi_{\varepsilon_n}[P|s]$ concentrates onto P^*,

Preliminary Theoretical Results

- So $\Pi_{\varepsilon_n}[P|s]$ concentrates onto P^*,

Preliminary Theoretical Results

- So $\Pi_{\varepsilon_n}[P|s]$ concentrates onto P^*, with P^* determined by the choice of loss function,

Preliminary Theoretical Results

- So $\Pi_{\varepsilon_n}[P|s]$ concentrates onto P^*, with P^* determined by the choice of loss function, the choice of $\mathcal{P}$,

Preliminary Theoretical Results

- So $\Pi_{\varepsilon_n}[P|s]$ concentrates onto P^*, with P^* determined by the choice of loss function, the choice of $\mathcal{P}$, and by the **true** $F[.|y_{1:t}]$

Preliminary Theoretical Results

- So $\Pi_{\varepsilon_n}[P|s]$ concentrates onto P^*, with P^* determined by the choice of loss function, the choice of $\mathcal{P}$, and by the **true** $F[.|y_{1:t}]$
- How does $\Pi_{\varepsilon_n}[P|s]$ relate to the true $F[.|y_{1:t}]$?

Preliminary Theoretical Results

- So $\Pi_{\varepsilon_n}[P|s]$ concentrates onto P^*, with P^* determined by the choice of loss function, the choice of $\mathcal{P}$, and by the **true** $F[.|y_{1:t}]$
- How does $\Pi_{\varepsilon_n}[P|s]$ relate to the true $F[.|y_{1:t}]$?
- Define:

$$\begin{aligned} \overline{P}_{\varepsilon_n}[.|y_{1:t}] &= \int_{\mathcal{P}} P[.|y_{1:t}] d\Pi_{\varepsilon_n}[P|s] \\ &= \text{the posterior mean of } P \end{aligned}$$

Preliminary Theoretical Results

- **Theorem 2: Predictive Merging.** As $n \longrightarrow \infty$ and $\varepsilon_n \longrightarrow \varepsilon^*$

- **Theorem 2: Predictive Merging.** As $n \to \infty$ and $\varepsilon_n \to \varepsilon^*$

(a) If $F \in \mathcal{P}$ (i.e. when the **true predictive** is in the class) we **do recover it**:

$$\rho^2_{TV}\left(\overline{P}_{\varepsilon_n}[.|y_{1:t}], F[.|y_{1:t}]\right) \quad \to \quad 0$$

Preliminary Theoretical Results

- **Theorem 2: Predictive Merging.** As $n \to \infty$ and $\varepsilon_n \to \varepsilon^*$

(a) If $F \in \mathcal{P}$ (i.e. when the **true predictive** is in the class) we **do recover it**:

$$\rho^2_{TV}\left(\overline{P}_{\varepsilon_n}[.|y_{1:t}], F[.|y_{1:t}]\right) \quad \to \quad 0$$

- i.e. (squared) total variation distance of $\overline{P}_{\varepsilon_n}[.|y_{1:t}]$ from the true predictive $\to 0$

Preliminary Theoretical Results

- **Theorem 2: Predictive Merging.** As $n \longrightarrow \infty$ and $\varepsilon_n \longrightarrow \varepsilon^*$

Preliminary Theoretical Results

- **Theorem 2: Predictive Merging.** As $n \longrightarrow \infty$ and $\varepsilon_n \longrightarrow \varepsilon^*$

(b) If $F \notin \mathcal{P}$ (so under **mis-specification**):

$$\lim_{n\to\infty} \rho^2_{TV}\left(\overline{P}_{\varepsilon_n}[.|y_{1:t}], F[.|y_{1:t}]\right) \leq 4\rho^2_{Hellinger}(P^*, F)$$

Preliminary Theoretical Results

- **Theorem 2: Predictive Merging.** As $n \to \infty$ and $\varepsilon_n \to \varepsilon^*$

(b) If $F \notin \mathcal{P}$ (so under **mis-specification**):

$$\lim_{n\to\infty} \rho^2_{TV}\left(\overline{P}_{\varepsilon_n}[.|y_{1:t}], F[.|y_{1:t}]\right) \leq 4\rho^2_{Hellinger}\left(P^*, F\right)$$

- P^* = predictive distribution that **maximizes the expected score** $\Leftrightarrow$ **is closest to** F in this sense

Preliminary Theoretical Results

- **Theorem 2: Predictive Merging.** As $n \to \infty$ and $\varepsilon_n \to \varepsilon^*$

(b) If $F \notin \mathcal{P}$ (so under **mis-specification**):

$$\lim_{n\to\infty} \rho^2_{TV}\left(\overline{P}_{\varepsilon_n}[.|y_{1:t}], F[.|y_{1:t}]\right) \leq 4\rho^2_{Hellinger}(P^*, F)$$

- $P^* =$ predictive distribution that **maximizes the expected score** $\Leftrightarrow$ **is closest to** F in this sense
- $\Rightarrow$ for a **given** class $P \in \mathcal{P}$, and **given** score (or loss) the bound is **the tightest possible**

Preliminary Theoretical Results

- **Theorem 2: Predictive Merging.** As $n \rightarrow \infty$ and $\varepsilon_n \rightarrow \varepsilon^*$

(b) If $F \notin \mathcal{P}$ (so under **mis-specification**):

$$\lim_{n\rightarrow\infty} \rho^2_{TV}\left(\overline{P}_{\varepsilon_n}[.|y_{1:t}], F[.|y_{1:t}]\right) \leq 4\rho^2_{Hellinger}(P^*, F)$$

- P^* = predictive distribution that **maximizes the expected score** $\Leftrightarrow$ **is closest to** F in this sense
- $\Rightarrow$ for a **given** class $P \in \mathcal{P}$, and **given** score (or loss) the bound is **the tightest possible**
- **Actual magnitude** of the bound is (of course) affected by $\mathcal{P}$ and the chosen loss function

Illustrative Example: Financial Asset Return

Illustrative Example: Financial Asset Return

- Let $\ln S_t$ = log of an asset price

Illustrative Example: Financial Asset Return

- Let $\ln S_t$ = log of an asset price

Illustrative Example: Financial Asset Return

- Let $\ln S_t$ = log of an asset price
- Let $\mathcal{P}$ define a class of predictives, $P_{\boldsymbol{\theta}}$,

Illustrative Example: Financial Asset Return

- Let $\ln S_t =$ log of an asset price
- Let $\mathcal{P}$ define a class of predictives, $P_{\boldsymbol{\theta}}$,

Illustrative Example: Financial Asset Return

- Let $\ln S_t =$ log of an asset price
- Let $\mathcal{P}$ define a class of predictives, $P_{\boldsymbol{\theta}}$, associated with a **stochastic volatility** model

$$
\begin{aligned}
d \ln S_t &= \sqrt{V_t} dB_t^S \\
dV_t &= (\theta_1 - \theta_2 V_t)\, dt + \theta_3 \sqrt{V_t} dB_t^V
\end{aligned}
$$

Illustrative Example: Financial Asset Return

- Let $\ln S_t$ = log of an asset price
- Let $\mathcal{P}$ define a class of predictives, $P_{\boldsymbol{\theta}}$, associated with a **stochastic volatility** model

$$d \ln S_t = \sqrt{V_t} dB_t^S$$
$$dV_t = (\theta_1 - \theta_2 V_t)\, dt + \theta_3 \sqrt{V_t} dB_t^V$$

- with $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$

Illustrative Example: Financial Asset Return

- Let $\ln S_t$ = log of an asset price
- Let $\mathcal{P}$ define a class of predictives, $P_{\boldsymbol{\theta}}$, associated with a **stochastic volatility** model

$$
\begin{aligned}
d\ln S_t &= \sqrt{V_t}dB_t^S \\
dV_t &= (\theta_1 - \theta_2 V_t)\,dt + \theta_3\sqrt{V_t}dB_t^V
\end{aligned}
$$

- with $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$
- The **true DGP**, F, is a stochastic volatility model with random **jumps**:

$$
\begin{aligned}
d\ln S_t &= \sqrt{V_t}dB_t^S + \underbrace{Z_t dN_t}_{=g(\theta_{0,4},\theta_{0,5}\dots)} \\
dV_t &= (\theta_{0,1} - \theta_{0,2} V_t)\,dt + \theta_{0,3}\sqrt{V_t}dB_t^V
\end{aligned}
$$

Illustrative Example: Financial Asset Return

- Let $\ln S_t$ = log of an asset price
- Let $\mathcal{P}$ define a class of predictives, $P_{\boldsymbol{\theta}}$, associated with a **stochastic volatility** model

$$d \ln S_t = \sqrt{V_t} dB_t^S$$
$$dV_t = (\theta_1 - \theta_2 V_t)\, dt + \theta_3 \sqrt{V_t} dB_t^V$$

- with $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$
- The **true DGP**, F, is a stochastic volatility model with random **jumps**:

$$d \ln S_t = \sqrt{V_t} dB_t^S + \underbrace{Z_t dN_t}_{= g(\theta_{0,4}, \theta_{0,5} \dots)}$$
$$dV_t = (\theta_{0,1} - \theta_{0,2} V_t)\, dt + \theta_{0,3} \sqrt{V_t} dB_t^V$$

  - $\boldsymbol{\theta}_0 = (\theta_{0,1}, \theta_{0,2}, \theta_{0,3}, \dots)' =$ **true parameter** (vector)

Illustrative Example

- If we **were** to simply adopt the (implied) **mis-specified SV** model for

$$y_t = \ln S_t - \ln S_{t-1} = \textbf{return} \text{ at time } t$$

Illustrative Example

- If we **were** to simply adopt the (implied) **mis-specified SV** model for

$$y_t = \ln S_t - \ln S_{t-1} = \textbf{return} \text{ at time } t$$

- and produce the conventional predictive: $p(y_{T+1}|y_{1:T})$

Illustrative Example

- If we **were** to simply adopt the (implied) **mis-specified SV** model for

$$y_t = \ln S_t - \ln S_{t-1} = \textbf{return} \text{ at time } t$$

- and produce the conventional predictive: $p(y_{T+1}|y_{1:T})$
- What would we find?

Illustrative Example

- If we **were** to simply adopt the (implied) **mis-specified SV** model for

$$y_t = \ln S_t - \ln S_{t-1} = \textbf{return} \text{ at time } t$$

- and produce the conventional predictive: $p(y_{T+1}|y_{1:T})$
- What would we find?
- $p(\boldsymbol{\theta}|y_{1:T})$ (under regul.) concentrates onto **pseudo-true** $\boldsymbol{\theta}$, $\boldsymbol{\theta}^*$

Illustrative Example

- If we **were** to simply adopt the (implied) **mis-specified SV** model for

$$y_t = \ln S_t - \ln S_{t-1} = \textbf{return} \text{ at time } t$$

- and produce the conventional predictive: $p(y_{T+1}|y_{1:T})$
- What would we find?
- $p(\boldsymbol{\theta}|y_{1:T})$ (under regul.) concentrates onto **pseudo-true** $\boldsymbol{\theta}$, $\boldsymbol{\theta}^*$
- where $\boldsymbol{\theta}^*$ is close to $\boldsymbol{\theta}_0$ (in KL-based sense)

Illustrative Example

- If we **were** to simply adopt the (implied) **mis-specified SV** model for

$$y_t = \ln S_t - \ln S_{t-1} = \textbf{return} \text{ at time } t$$

- and produce the conventional predictive: $p(y_{T+1}|y_{1:T})$
- What would we find?
- $p(\boldsymbol{\theta}|y_{1:T})$ (under regul.) concentrates onto **pseudo-true** $\boldsymbol{\theta}$, $\boldsymbol{\theta}^*$
- where $\boldsymbol{\theta}^*$ is close to $\boldsymbol{\theta}_0$ (in KL-based sense)
- $\Rightarrow$

$$\lim_{T\rightarrow\infty} p(y_{T+1}|y_{1:T}) = p(y_{T+1}|y_{1:T}, \boldsymbol{\theta}^*)$$

Illustrative Example

- If we **were** to simply adopt the (implied) **mis-specified SV** model for

$$y_t = \ln S_t - \ln S_{t-1} = \textbf{return} \text{ at time } t$$

- and produce the conventional predictive: $p(y_{T+1}|y_{1:T})$
- What would we find?
- $p(\boldsymbol{\theta}|y_{1:T})$ (under regul.) concentrates onto **pseudo-true** $\boldsymbol{\theta}$, $\boldsymbol{\theta}^*$
- where $\boldsymbol{\theta}^*$ is close to $\boldsymbol{\theta}_0$ (in KL-based sense)
- $\Rightarrow$

$$\lim_{T\to\infty} p(y_{T+1}|y_{1:T}) = p(y_{T+1}|y_{1:T}, \boldsymbol{\theta}^*)$$

Illustrative Example

- If we **were** to simply adopt the (implied) **mis-specified SV** model for

$$y_t = \ln S_t - \ln S_{t-1} = \textbf{return} \text{ at time } t$$

- and produce the conventional predictive: $p(y_{T+1}|y_{1:T})$
- What would we find?
- $p(\boldsymbol{\theta}|y_{1:T})$ (under regul.) concentrates onto **pseudo-true** $\boldsymbol{\theta}$, $\boldsymbol{\theta}^*$
- where $\boldsymbol{\theta}^*$ is close to $\boldsymbol{\theta}_0$ (in KL-based sense)
- $\Rightarrow$

$$\lim_{T\to\infty} p(y_{T+1}|y_{1:T}) = p(y_{T+1}|y_{1:T}, \boldsymbol{\theta}^*) = \mathit{what??}$$

Illustrative Example

- If we **were** to simply adopt the (implied) **mis-specified SV** model for

$$y_t = \ln S_t - \ln S_{t-1} = \textbf{return} \text{ at time } t$$

- and produce the conventional predictive: $p(y_{T+1}|y_{1:T})$
- What would we find?
- $p(\boldsymbol{\theta}|y_{1:T})$ (under regul.) concentrates onto **pseudo-true** $\boldsymbol{\theta}$, $\boldsymbol{\theta}^*$
- where $\boldsymbol{\theta}^*$ is close to $\boldsymbol{\theta}_0$ (in KL-based sense)
- $\Rightarrow$

$$\lim_{T\to\infty} p(y_{T+1}|y_{1:T}) = p(y_{T+1}|y_{1:T}, \boldsymbol{\theta}^*) = \mathit{what??}$$

Illustrative Example

- P is misspecified

Illustrative Example

- P is misspecified
- $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$

Illustrative Example

- P is misspecified
- $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$
- **Minimizing** KL divergence $\equiv$ **maximizing log score** *in sample*

Illustrative Example

- P is misspecified
- $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$
- **Minimizing** KL divergence $\equiv$ **maximizing log score** *in sample*
- **No guarantee** of *out-of-sample* performance

Illustrative Example

- P is misspecified
- $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$
- **Minimizing** KL divergence $\equiv$ **maximizing log score** *in sample*
- **No guarantee** of *out-of-sample* performance
- **FBF ensures** accurate *out-of-sample* performance according to any given score/loss

Focused Bayesian Prediction

Focused Bayesian Prediction

- **Five** loss functions considered:

Focused Bayesian Prediction

- **Five** loss functions considered:
  - **Three scores:**

Focused Bayesian Prediction

- **Five** loss functions considered:
  - **Three scores:**
    1. Log score

Focused Bayesian Prediction

- **Five** loss functions considered:
  - **Three scores:**
    1. Log score
    2. Continuous rank probability score (CRPS)

Focused Bayesian Prediction

- **Five** loss functions considered:
  - **Three scores:**
    1. Log score
    2. Continuous rank probability score (CRPS)
    3. CRPS for lower tail (appropriate for a financial return)

Focused Bayesian Prediction

- **Five** loss functions considered:
  - **Three scores:**
    1. Log score
    2. Continuous rank probability score (CRPS)
    3. CRPS for lower tail (appropriate for a financial return)
  - **Two auxiliary predictive**-based losses

Focused Bayesian Prediction

- **Five** loss functions considered:
  - **Three scores:**
    1. Log score
    2. Continuous rank probability score (CRPS)
    3. CRPS for lower tail (appropriate for a financial return)
  - **Two auxiliary predictive**-based losses
  - Adopting the flavour of **auxiliary model-based** ABC

Auxiliary model-based ABC

- **Drovandi et al. (2011, 2015, 2018); Creel and Kristensen (2015); Drovandi (2018); Martin et al. (2018)**

Auxiliary model-based ABC

- **Drovandi et al. (2011, 2015, 2018); Creel and Kristensen (2015); Drovandi (2018); Martin et al. (2018)**
- Specify a **tractable** $q(y_{1:T}|\boldsymbol{\beta})$ that *approximates* $p(y_{1:T}|\boldsymbol{\theta})$

Auxiliary model-based ABC

- **Drovandi et al. (2011, 2015, 2018); Creel and Kristensen (2015); Drovandi (2018); Martin et al. (2018)**
- Specify a **tractable** $q(y_{1:T}|\boldsymbol{\beta})$ that *approximates* $p(y_{1:T}|\boldsymbol{\theta})$
- $\widehat{\boldsymbol{\beta}}_{MLE} \Rightarrow$ 'summary statistic' $\boldsymbol{\eta}(y_{1:T})$

Auxiliary model-based ABC

- **Drovandi et al. (2011, 2015, 2018); Creel and Kristensen (2015); Drovandi (2018); Martin et al. (2018)**
- Specify a **tractable** $q(y_{1:T}|\boldsymbol{\beta})$ that *approximates* $p(y_{1:T}|\boldsymbol{\theta})$
- $\widehat{\boldsymbol{\beta}}_{MLE} \Rightarrow$ 'summary statistic' $\boldsymbol{\eta}(y_{1:T})$
- Produce **approximate posterior**, $p(\boldsymbol{\theta}|\boldsymbol{\eta}(y_{1:T}))$

Auxiliary model-based ABC

- **Drovandi et al. (2011, 2015, 2018); Creel and Kristensen (2015); Drovandi (2018); Martin et al. (2018)**
- Specify a **tractable** $q(y_{1:T}|\boldsymbol{\beta})$ that *approximates* $p(y_{1:T}|\boldsymbol{\theta})$
- $\widehat{\boldsymbol{\beta}}_{MLE} \Rightarrow$ 'summary statistic' $\boldsymbol{\eta}(y_{1:T})$
- Produce **approximate posterior**, $p(\boldsymbol{\theta}|\boldsymbol{\eta}(y_{1:T}))$
- Aim in auxiliary model-based ABC?

Auxiliary model-based ABC

- **Drovandi et al. (2011, 2015, 2018); Creel and Kristensen (2015); Drovandi (2018); Martin et al. (2018)**
- Specify a **tractable** $q(y_{1:T}|\boldsymbol{\beta})$ that *approximates* $p(y_{1:T}|\boldsymbol{\theta})$
- $\widehat{\boldsymbol{\beta}}_{MLE} \Rightarrow$ 'summary statistic' $\boldsymbol{\eta}(y_{1:T})$
- Produce **approximate posterior**, $p(\boldsymbol{\theta}|\boldsymbol{\eta}(y_{1:T}))$
- Aim in auxiliary model-based ABC?
- Choose $q(y_{1:T}|\boldsymbol{\beta})$ to capture features of $p(y_{1:T}|\boldsymbol{\theta})$

Auxiliary model-based ABC

- **Drovandi et al. (2011, 2015, 2018); Creel and Kristensen (2015); Drovandi (2018); Martin et al. (2018)**
- Specify a **tractable** $q(y_{1:T}|\boldsymbol{\beta})$ that *approximates* $p(y_{1:T}|\boldsymbol{\theta})$
- $\widehat{\boldsymbol{\beta}}_{MLE} \Rightarrow$ 'summary statistic' $\boldsymbol{\eta}(y_{1:T})$
- Produce **approximate posterior**, $p(\boldsymbol{\theta}|\boldsymbol{\eta}(y_{1:T}))$
- Aim in auxiliary model-based ABC?
- Choose $q(y_{1:T}|\boldsymbol{\beta})$ to capture features of $p(y_{1:T}|\boldsymbol{\theta})$
- If $q(y_{1:T}|\boldsymbol{\beta})$ 'nests' (a **correctly specified**) $p(y_{1:T}|\boldsymbol{\theta})$

Auxiliary model-based ABC

- **Drovandi et al. (2011, 2015, 2018); Creel and Kristensen (2015); Drovandi (2018); Martin et al. (2018)**
- Specify a **tractable** $q(y_{1:T}|\boldsymbol{\beta})$ that *approximates* $p(y_{1:T}|\boldsymbol{\theta})$
- $\widehat{\boldsymbol{\beta}}_{MLE} \Rightarrow$ 'summary statistic' $\boldsymbol{\eta}(y_{1:T})$
- Produce **approximate posterior**, $p(\boldsymbol{\theta}|\boldsymbol{\eta}(y_{1:T}))$
- Aim in auxiliary model-based ABC?
- Choose $q(y_{1:T}|\boldsymbol{\beta})$ to capture features of $p(y_{1:T}|\boldsymbol{\theta})$
- If $q(y_{1:T}|\boldsymbol{\beta})$ 'nests' (a **correctly specified**) $p(y_{1:T}|\boldsymbol{\theta})$
    - $\Rightarrow \boldsymbol{\eta}(y_{1:T}) = \widehat{\boldsymbol{\beta}}_{MLE}$ is **asymptotically sufficient** for $\boldsymbol{\theta}$

Auxiliary model-based ABC

- **Drovandi et al. (2011, 2015, 2018); Creel and Kristensen (2015); Drovandi (2018); Martin et al. (2018)**
- Specify a **tractable** $q(y_{1:T}|\boldsymbol{\beta})$ that *approximates* $p(y_{1:T}|\boldsymbol{\theta})$
- $\widehat{\boldsymbol{\beta}}_{MLE} \Rightarrow$ 'summary statistic' $\boldsymbol{\eta}(y_{1:T})$
- Produce **approximate posterior**, $p(\boldsymbol{\theta}|\boldsymbol{\eta}(y_{1:T}))$
- Aim in auxiliary model-based ABC?
- Choose $q(y_{1:T}|\boldsymbol{\beta})$ to capture features of $p(y_{1:T}|\boldsymbol{\theta})$
- If $q(y_{1:T}|\boldsymbol{\beta})$ 'nests' (a **correctly specified**) $p(y_{1:T}|\boldsymbol{\theta})$
  - $\Rightarrow \boldsymbol{\eta}(y_{1:T}) = \widehat{\boldsymbol{\beta}}_{MLE}$ is **asymptotically sufficient** for $\boldsymbol{\theta}$
  - $\Rightarrow p(\boldsymbol{\theta}|\boldsymbol{\eta}(y_{1:T})) = p(\boldsymbol{\theta}|y_{1:T})$ (for large T)

Auxiliary model-based ABC

- **Drovandi et al. (2011, 2015, 2018); Creel and Kristensen (2015); Drovandi (2018); Martin et al. (2018)**
- Specify a **tractable** $q(y_{1:T}|\boldsymbol{\beta})$ that *approximates* $p(y_{1:T}|\boldsymbol{\theta})$
- $\widehat{\boldsymbol{\beta}}_{MLE} \Rightarrow$ 'summary statistic' $\boldsymbol{\eta}(y_{1:T})$
- Produce **approximate posterior**, $p(\boldsymbol{\theta}|\boldsymbol{\eta}(y_{1:T}))$
- Aim in auxiliary model-based ABC?
- Choose $q(y_{1:T}|\boldsymbol{\beta})$ to capture features of $p(y_{1:T}|\boldsymbol{\theta})$
- If $q(y_{1:T}|\boldsymbol{\beta})$ 'nests' (a **correctly specified**) $p(y_{1:T}|\boldsymbol{\theta})$
    - $\Rightarrow \boldsymbol{\eta}(y_{1:T}) = \widehat{\boldsymbol{\beta}}_{MLE}$ is **asymptotically sufficient** for $\boldsymbol{\theta}$
    - $\Rightarrow p(\boldsymbol{\theta}|\boldsymbol{\eta}(y_{1:T})) = p(\boldsymbol{\theta}|y_{1:T})$ (for large T)
    - $\Rightarrow$ 'ideal' $q(y_{1:T}|\boldsymbol{\beta})$ is **highly parameterized**

Auxiliary predictive-based loss function

- But that do we know about **prediction**??

Auxiliary predictive-based loss function

- But that do we know about **prediction**??
- **Simple parsimoneous** models often forecast better than **complex, highly parameterized (but incorrect)** models....

Auxiliary predictive-based loss function

- But that do we know about **prediction**??
- **Simple parsimoneous** models often forecast better than **complex, highly parameterized (but incorrect)** models....
- $\Rightarrow$ Approach in auxiliary-model based **focused Bayesian prediction**?

Auxiliary predictive-based loss function

- But that do we know about **prediction**??
- **Simple parsimoneous** models often forecast better than **complex, highly parameterized (but incorrect)** models....
- $\Rightarrow$ Approach in auxiliary-model based **focused Bayesian prediction**?
- Pick a **simple parsimoneous 'auxiliary predictive':** $q(y_{t+1}|y_{1:t}, \beta)$

Auxiliary predictive-based loss function

- But that do we know about **prediction**??
- **Simple parsimoneous** models often forecast better than **complex, highly parameterized (but incorrect)** models....
- $\Rightarrow$ Approach in auxiliary-model based **focused Bayesian prediction**?
- Pick a **simple parsimoneous 'auxiliary predictive':** $q(y_{t+1}|y_{1:t}, \boldsymbol{\beta})$
- And **select** $p(y_{t+1}|y_{1:t}, \boldsymbol{\theta}^i)$ such that its predictive performance closely **matches** that of $q(y_{t+1}|y_{1:t}, \boldsymbol{\beta})$ over the test period

Auxiliary predictive-based loss function

- i.e. **select** $p(y_{t+1}|y_{1:t},\boldsymbol{\theta}^i)$ such that:

$$\frac{1}{n}\sum_{i=0}^{n-1}\left|p(y_{(\tau+i)+1}|y_{1:(\tau+i)},\boldsymbol{\theta}^i) - q(y_{(\tau+i)+1}|y_{1:(\tau+i)},\widehat{\boldsymbol{\beta}})\right|$$

$<$ the **lowest** (α%, say) quantile

Auxiliary predictive-based loss function

- i.e. **select** $p(y_{t+1}|y_{1:t}, \boldsymbol{\theta}^i)$ such that:

$$\frac{1}{n}\sum_{i=0}^{n-1}\left|p(y_{(\tau+i)+1}|y_{1:(\tau+i)}, \boldsymbol{\theta}^i) - q(y_{(\tau+i)+1}|y_{1:(\tau+i)}, \widehat{\boldsymbol{\beta}})\right|$$

$<$ the **lowest** (α%, say) quantile

- i.e. such that **loss** (defined by this predictive difference) is **small**

Auxiliary predictive-based loss function

- i.e. **select** $p(y_{t+1}|y_{1:t}, \boldsymbol{\theta}^i)$ such that:

$$\frac{1}{n}\sum_{i=0}^{n-1}\left|p(y_{(\tau+i)+1}|y_{1:(\tau+i)}, \boldsymbol{\theta}^i) - q(y_{(\tau+i)+1}|y_{1:(\tau+i)}, \widehat{\boldsymbol{\beta}})\right|$$

$<$ the **lowest** (α%, say) quantile

- i.e. such that **loss** (defined by this predictive difference) is **small**
- Choose $q(y_{t+1}|y_{1:t}, \boldsymbol{\beta})$ to be a **generalized autoregressive conditionally heteroscedastic (GARCH)** model

Auxiliary predictive-based loss function

- i.e. **select** $p(y_{t+1}|y_{1:t}, \theta^i)$ such that:

$$\frac{1}{n}\sum_{i=0}^{n-1}\left|p(y_{(\tau+i)+1}|y_{1:(\tau+i)}, \theta^i) - q(y_{(\tau+i)+1}|y_{1:(\tau+i)}, \widehat{\beta})\right|$$

$<$ the **lowest** (α%, say) quantile

- i.e. such that **loss** (defined by this predictive difference) is **small**
- Choose $q(y_{t+1}|y_{1:t}, \beta)$ to be a **generalized autoregressive conditionally heteroscedastic (GARCH)** model
  - with Student t errors (work-horse of empirical finance)

Auxiliary predictive-based loss function

- i.e. **select** $p(y_{t+1}|y_{1:t}, \boldsymbol{\theta}^i)$ such that:

$$\frac{1}{n}\sum_{i=0}^{n-1}\left|p(y_{(\tau+i)+1}|y_{1:(\tau+i)}, \boldsymbol{\theta}^i) - q(y_{(\tau+i)+1}|y_{1:(\tau+i)}, \widehat{\boldsymbol{\beta}})\right|$$

$<$ the **lowest** (α%, say) quantile

- i.e. such that **loss** (defined by this predictive difference) is **small**
- Choose $q(y_{t+1}|y_{1:t}, \boldsymbol{\beta})$ to be a **generalized autoregressive conditionally heteroscedastic (GARCH)** model
  - with Student t errors (work-horse of empirical finance)
  - with normal errors (expected to be a poorer 'benchmark')

Numerical results

Numerical results

- Display draws from each $\Pi_{\varepsilon_n}[P|s]$

Numerical results

- Display draws from each $\Pi_{\varepsilon_n}[P|s]$
- (associated with one particular out-of-sample time period)

Numerical results

- Display draws from each $\Pi_{\varepsilon_n}[P|s]$
- (associated with one particular out-of-sample time period)
- (Using the notation s for all five posteriors)

Numerical results

- Display draws from each $\Pi_{\varepsilon_n}[P|s]$
- (associated with one particular out-of-sample time period)
- (Using the notation s for all five posteriors)
- Estimate: $\overline{P}_{\varepsilon_n}[.|y_{1:t}] = \int_{\mathcal{P}} P[.|y_{1:t}] d\Pi_{\varepsilon_n}[P|s]$

Numerical results

- Display draws from each $\Pi_{\varepsilon_n}[P|s]$
- (associated with one particular out-of-sample time period)
- (Using the notation s for all five posteriors)
- Estimate: $\overline{P}_{\varepsilon_n}[.|y_{1:t}] = \int_{\mathcal{P}} P[.|y_{1:t}] d\Pi_{\varepsilon_n}[P|s]$
- Roll the whole process forward:

Numerical results

- Display draws from each $\Pi_{\varepsilon_n}[P|s]$
- (associated with one particular out-of-sample time period)
- (Using the notation s for all five posteriors)
- Estimate: $\overline{P}_{\varepsilon_n}[.|y_{1:t}] = \int_{\mathcal{P}} P[.|y_{1:t}] d\Pi_{\varepsilon_n}[P|s]$
- Roll the whole process forward:
- Compute (over 100 (truely) out-of-sample periods):

Numerical results

- Display draws from each $\Pi_{\varepsilon_n}[P|s]$
- (associated with one particular out-of-sample time period)
- (Using the notation s for all five posteriors)
- Estimate: $\overline{P}_{\varepsilon_n}[.|y_{1:t}] = \int_{\mathcal{P}} P[.|y_{1:t}]d\Pi_{\varepsilon_n}[P|s]$
- Roll the whole process forward:
- Compute (over 100 (truely) out-of-sample periods):
- Median:

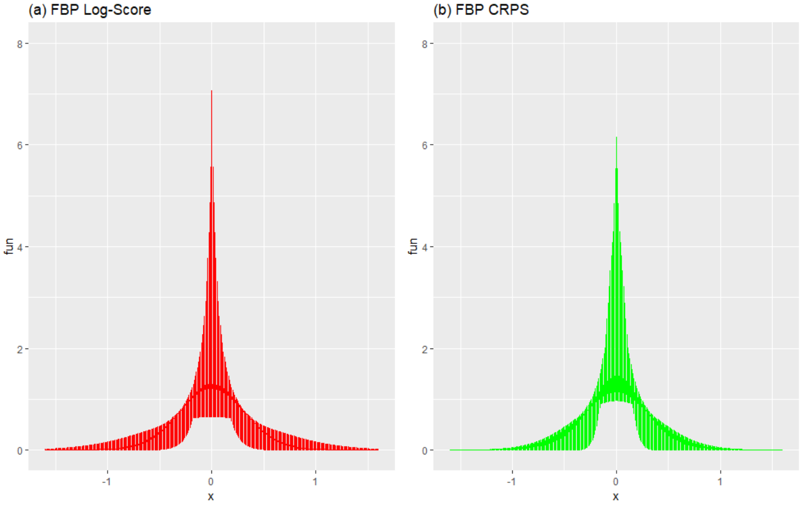
Numerical results

- Display draws from each $\Pi_{\varepsilon_n}[P|s]$
- (associated with one particular out-of-sample time period)
- (Using the notation s for all five posteriors)
- Estimate: $\overline{P}_{\varepsilon_n}[.|y_{1:t}] = \int_{\mathcal{P}} P[.|y_{1:t}] d\Pi_{\varepsilon_n}[P|s]$
- Roll the whole process forward:
- Compute (over 100 (truely) out-of-sample periods):
- Median:
  - **log scores; CRPS scores; tail-weighted CRPS scores**

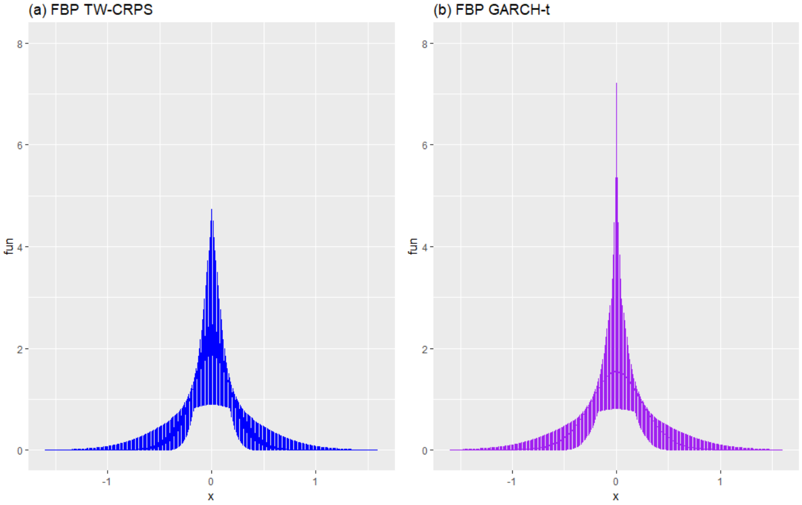
Numerical results

- Display draws from each $\Pi_{\varepsilon_n}[P|s]$
- (associated with one particular out-of-sample time period)
- (Using the notation s for all five posteriors)
- Estimate: $\overline{P}_{\varepsilon_n}[.|y_{1:t}] = \int_{\mathcal{P}} P[.|y_{1:t}] d\Pi_{\varepsilon_n}[P|s]$
- Roll the whole process forward:
- Compute (over 100 (truely) out-of-sample periods):
- Median:
    - **log scores; CRPS scores; tail-weighted CRPS scores**
- Compare with results for **exact (MCMC) mis-specified:** $p(y_{t+1}|y_{1:t})$

Posterior distributions



Posterior distributions



Median Scores (100 out-of-sample predictions)

Median Scores (100 out-of-sample predictions)

- The loss function based on matching the Student t GARCH (auxiliary) predictive **yields the most accurate predictive** - according to all measures of predictive accuracy

Median Scores (100 out-of-sample predictions)

- The loss function based on matching the Student t GARCH (auxiliary) predictive **yields the most accurate predictive** - according to all measures of predictive accuracy
- The loss function based on the (raw) CRPS score is **second best** - according to all measures of predictive accuracy

Median Scores (100 out-of-sample predictions)

- The loss function based on matching the Student t GARCH (auxiliary) predictive **yields the most accurate predictive** - according to all measures of predictive accuracy
- The loss function based on the (raw) CRPS score is **second best** - according to all measures of predictive accuracy
- The loss function based on matching the normal GARCH (auxiliary) predictive does not - as anticipated - perform well

Median Scores (100 out-of-sample predictions)

- The loss function based on matching the Student t GARCH (auxiliary) predictive **yields the most accurate predictive** - according to all measures of predictive accuracy
- The loss function based on the (raw) CRPS score is **second best** - according to all measures of predictive accuracy
- The loss function based on matching the normal GARCH (auxiliary) predictive does not - as anticipated - perform well
- The **exact but mis-specified** predictive is **in the lower half of the ranking** in all cases.....

Median Scores (100 out-of-sample predictions)

- The loss function based on matching the Student t GARCH (auxiliary) predictive **yields the most accurate predictive** - according to all measures of predictive accuracy
- The loss function based on the (raw) CRPS score is **second best** - according to all measures of predictive accuracy
- The loss function based on matching the normal GARCH (auxiliary) predictive does not - as anticipated - perform well
- The **exact but mis-specified** predictive is **in the lower half of the ranking** in all cases.....
- So we *are* gaining in terms of predictive accuracy via **FBP**

Median Scores (100 out-of-sample predictions)

- The loss function based on matching the Student t GARCH (auxiliary) predictive **yields the most accurate predictive** - according to all measures of predictive accuracy
- The loss function based on the (raw) CRPS score is **second best** - according to all measures of predictive accuracy
- The loss function based on matching the normal GARCH (auxiliary) predictive does not - as anticipated - perform well
- The **exact but mis-specified** predictive is **in the lower half of the ranking** in all cases.....
- So we *are* gaining in terms of predictive accuracy via **FBP**
- A larger number of out-of-sample evaluations is needed for precise conclusions.....(the particle filtering takes time.....)

In Conclusion

- New loss-based approach to Bayesian forecasting

In Conclusion

- New loss-based approach to Bayesian forecasting
- Appropriate for an $\mathcal{M}$-open world

In Conclusion

- New loss-based approach to Bayesian forecasting
- Appropriate for an $\mathcal{M}$-open world
- (in which model-mis-specification is explicitly acknowledged)

In Conclusion

- New loss-based approach to Bayesian forecasting
- Appropriate for an $\mathcal{M}$-open world
- (in which model-mis-specification is explicitly acknowledged)
- **Focused Bayesian prediction** does improve predictive performance **under model mis-specification**

In Conclusion

- New loss-based approach to Bayesian forecasting
- Appropriate for an $\mathcal{M}$-open world
- (in which model-mis-specification is explicitly acknowledged)
- **Focused Bayesian prediction** does improve predictive performance **under model mis-specification**
- Relative to an exact (but mis-specified) predictive

In Conclusion

- New loss-based approach to Bayesian forecasting
- Appropriate for an $\mathcal{M}$-open world
- (in which model-mis-specification is explicitly acknowledged)
- **Focused Bayesian prediction** does improve predictive performance **under model mis-specification**
- Relative to an exact (but mis-specified) predictive
- (At least based on this small numerical exercise....)

In Conclusion

- New loss-based approach to Bayesian forecasting
- Appropriate for an $\mathcal{M}$-open world
- (in which model-mis-specification is explicitly acknowledged)
- **Focused Bayesian prediction** does improve predictive performance **under model mis-specification**
- Relative to an exact (but mis-specified) predictive
- (At least based on this small numerical exercise....)
- Important shift away from **Bissiri, Holmes and Walker (2016)**:

In Conclusion

- New loss-based approach to Bayesian forecasting
- Appropriate for an $\mathcal{M}$-open world
- (in which model-mis-specification is explicitly acknowledged)
- **Focused Bayesian prediction** does improve predictive performance **under model mis-specification**
- Relative to an exact (but mis-specified) predictive
- (At least based on this small numerical exercise....)
- Important shift away from **Bissiri, Holmes and Walker (2016)**:
- "*A general framework for updating belief distributions*"

In Conclusion

- Via **BHW**: **could** specify loss via a particular $S_n(P, F)$

In Conclusion

- Via **BHW**: **could** specify loss via a particular $S_n(P, F)$
- And update prior beliefs $\pi(P)$ as:

$$\pi[P|s] \propto \exp\left[-nS_n(P, F)\right]\pi[P]$$

In Conclusion

- Via **BHW**: **could** specify loss via a particular $S_n(P, F)$
- And update prior beliefs $\pi(P)$ as:

$$\pi[P|s] \propto \exp\left[-nS_n(P, F)\right] \pi[P]$$

- $\Rightarrow$ effectively assumes a **particular model** for $S_n(P, F)$:

$$g_n(s|P) \propto \exp[-nS_n(P, F)]$$

In Conclusion

- Via **BHW**: **could** specify loss via a particular $S_n(P, F)$
- And update prior beliefs $\pi(P)$ as:

$$\pi[P|s] \propto \exp\left[-nS_n(P, F)\right] \pi[P]$$

- $\Rightarrow$ effectively assumes a **particular model** for $S_n(P, F)$:

$$g_n(s|P) \propto \exp[-nS_n(P, F)]$$

- We make no such assumption $\Rightarrow$ allow the data to detemine $g_n(s|P)$

In Conclusion

- Via **BHW**: **could** specify loss via a particular $S_n(P, F)$
- And update prior beliefs $\pi(P)$ as:

$$\pi[P|s] \propto \exp\left[-nS_n(P, F)\right]\pi[P]$$

- $\Rightarrow$ effectively assumes a **particular model** for $S_n(P, F)$:

$$g_n(s|P) \propto \exp[-nS_n(P, F)]$$

- We make no such assumption $\Rightarrow$ allow the data to detemine $g_n(s|P)$
- With predictive accuracy guaranteed asymptotically

In Conclusion

- Via **BHW**: **could** specify loss via a particular $S_n(P, F)$
- And update prior beliefs $\pi(P)$ as:

$$\pi[P|s] \propto \exp\left[-nS_n(P, F)\right] \pi[P]$$

- $\Rightarrow$ effectively assumes a **particular model** for $S_n(P, F)$:

$$g_n(s|P) \propto \exp[-nS_n(P, F)]$$

- We make no such assumption $\Rightarrow$ allow the data to detemine $g_n(s|P)$
- With predictive accuracy guaranteed asymptotically
- More to come.....