Approximate Bayesian Forecasting in Intractable and Misspecified State Space Models

Chaya Weerasinghe

Department of Econometrics and Business Statistics
Monash University, Australia

Bayes Comp, March, 2023

Based on joint work with:
Professor Gael M. Martin, Dr. Ruben Loaiza Maya and Associate Professor David Frazier

State Space Models

State Space Models

- Powerful tool for modelling and forecasting the complex dynamics associated with the observed time series data

State Space Models

- Powerful tool for modelling and forecasting the complex dynamics associated with the observed time series data
- Observed data: $y_{1:T} = \{y_1, y_2 ..., y_T\}$

State Space Models

- Powerful tool for modelling and forecasting the complex dynamics associated with the observed time series data
- Observed data: $y_{1:T} = \{y_1, y_2 ..., y_T\}$
- Unobserved states: $x_{0:T} = \{x_0, x_2, ..., x_T\}$

State Space Models

- Powerful tool for modelling and forecasting the complex dynamics associated with the observed time series data
- Observed data: $y_{1:T} = \{y_1, y_2..., y_T\}$
- Unobserved states: $x_{0:T} = \{x_0, x_2, ..., x_T\}$
- A state space model (SSM) comprised of two parts:

State Space Models

- Powerful tool for modelling and forecasting the complex dynamics associated with the observed time series data
- Observed data: $y_{1:T} = \{y_1, y_2 ..., y_T\}$
- Unobserved states: $x_{0:T} = \{x_0, x_2, ..., x_T\}$
- A state space model (SSM) comprised of two parts:

State Space Models

- Powerful tool for modelling and forecasting the complex dynamics associated with the observed time series data
- Observed data: $y_{1:T} = \{y_1, y_2 ..., y_T\}$
- Unobserved states: $x_{0:T} = \{x_0, x_2, ..., x_T\}$
- A state space model (SSM) comprised of two parts:

Measurement density: $p(y_t|x_t, \theta)$

State Space Models

- Powerful tool for modelling and forecasting the complex dynamics associated with the observed time series data
- Observed data: $y_{1:T} = \{y_1, y_2..., y_T\}$
- Unobserved states: $x_{0:T} = \{x_0, x_2, ..., x_T\}$
- A state space model (SSM) comprised of two parts:

$$\text{Measurement density:} \qquad p(y_t|x_t, \theta)$$

$$\text{(Markov) Transition density:} \quad p(x_t|x_{t-1}, \theta)$$

State Space Models

- Powerful tool for modelling and forecasting the complex dynamics associated with the observed time series data
- Observed data: $y_{1:T} = \{y_1, y_2..., y_T\}$
- Unobserved states: $x_{0:T} = \{x_0, x_2, ..., x_T\}$
- A state space model (SSM) comprised of two parts:

$$\text{Measurement density:} \qquad p(y_t|x_t, \theta)$$

$$\text{(Markov) Transition density:} \quad p(x_t|x_{t-1}, \theta)$$

- Plus

$$\text{Initial state density:} \quad p(x_0|\theta)$$

State Space Models

- Powerful tool for modelling and forecasting the complex dynamics associated with the observed time series data
- Observed data: $y_{1:T} = \{y_1, y_2 ..., y_T\}$
- Unobserved states: $x_{0:T} = \{x_0, x_2, ..., x_T\}$
- A state space model (SSM) comprised of two parts:

$$\begin{aligned} &\text{Measurement density:} && p(y_t|x_t, \theta) \\ &\text{(Markov) Transition density:} && p(x_t|x_{t-1}, \theta) \end{aligned}$$

- Plus

$$\text{Initial state density:} \quad p(x_0|\theta)$$

- May depend on a set of unknown, **static** parameters θ

Exact Bayesian Forecasting in SSMs

- **Distribution of interest** is

Exact Bayesian Forecasting in SSMs

- **Distribution of interest** is

Exact Bayesian Forecasting in SSMs

- **Distribution of interest** is

$$p(y_{T+1}|y_{1:T})$$

Exact Bayesian Forecasting in SSMs

- **Distribution of interest** is

$$
\begin{aligned}
p(y_{T+1}|y_{1:T}) &= \int_{x_{T+1}} \int_{x_{0:T}} \int_{\theta} p(y_{T+1}|x_{T+1}, \theta, y_{1:T}) p(x_{T+1}|x_T, \theta) \\
&\quad \times \underbrace{\left(p(x_{0:T}|y_{1:T}, \theta)\right)} \underbrace{p\left(\theta|y_{1:T}\right)} d\theta dx_{0:T} dx_{T+1}
\end{aligned}
$$

Exact Bayesian Forecasting in SSMs

- **Distribution of interest** is

$$\begin{aligned} p(y_{T+1}|y_{1:T}) &= \int_{x_{T+1}} \int_{x_{0:T}} \int_{\theta} p(y_{T+1}|x_{T+1}, \theta, y_{1:T}) p(x_{T+1}|x_T, \theta) \\ &\quad \times \left(\underbrace{p(x_{0:T}|y_{1:T}, \theta)} \right) \underbrace{p\,(\theta|y_{1:T})}\, d\theta dx_{0:T} dx_{T+1} \end{aligned}$$

- Often readily accessible via
  - Bayesian Markov chain Monte Carlo (MCMC) methods or
  - Particle MCMC variants

Exact Bayesian Forecasting in SSMs

- **Distribution of interest** is

$$\begin{aligned} p(y_{T+1}|y_{1:T}) &= \int_{x_{T+1}} \int_{x_{0:T}} \int_{\theta} p(y_{T+1}|x_{T+1},\theta,y_{1:T}) p(x_{T+1}|x_T,\theta) \\ &\times \left(\underbrace{p(x_{0:T}|y_{1:T},\theta)}\right) \underbrace{p\left(\theta|y_{1:T}\right)} d\theta dx_{0:T} dx_{T+1} \end{aligned}$$

- Often readily accessible via
  - Bayesian Markov chain Monte Carlo (MCMC) methods or
  - Particle MCMC variants

- But... challenges remain when the model is **intractable** in some sense

Intractable SSMs

Intractable SSMs

- Occur in **two** ways:

Intractable SSMs

- Occur in **two** ways:

  1. when the **dimension** of either the $\mathbf{y}$ or the $\mathbf{x}$ (and the associated θ), or both, is **very large**

Intractable SSMs

- Occur in **two** ways:
  1. when the **dimension** of either the **y** or the **x** (and the associated θ), or both, is **very large**
  2. when the data generating process (DGP) **unavailable** in closed form

Intractable SSMs

- Occur in **two** ways:
  1. when the **dimension** of either the **y** or the **x** (and the associated θ), or both, is **very large**
  2. when the data generating process (DGP) **unavailable** in closed form
- $\Rightarrow$ exact Bayesian prediction **may not be feasible**

Intractable SSMs

Intractable SSMs

- Can be analyzed using ***approximate computational methods***

Intractable SSMs

- Can be analyzed using *approximate computational methods*
- High-dimensional SSMs are tackled via **variational Bayes** (VB) methods :
  - **[Tran et al., 2017], [Quiroz et al., 2022], [Loaiza-Maya et al., 2021b], [Frazier et al., 2022]**

Intractable SSMs

- Can be analyzed using *approximate computational methods*

- High-dimensional SSMs are tackled via **variational Bayes** (VB) methods :
  - **[Tran et al., 2017], [Quiroz et al., 2022], [Loaiza-Maya et al., 2021b], [Frazier et al., 2022]**

- SSMs with unavailable components are managed via **approximate Bayesian computation** (ABC) :
  - **[Dean et al., 2014], [Creel and Kristensen, 2015], [Frazier et al., 2019], [Martin et al., 2019]**

We focus on ABC...

- We explore the use of these approximate methods to conduct **forecasting** in intractable and ***misspecified*** SSMs

We focus on ABC...

- We explore the use of these approximate methods to conduct **forecasting** in intractable and ***misspecified*** SSMs
- with the aim of producing accurate, fit for purpose predictions

We focus on ABC...

- We explore the use of these approximate methods to conduct **forecasting** in intractable and ***misspecified*** SSMs

- with the aim of producing accurate, fit for purpose predictions

- We are explicitly looking at the **use of ABC**

We focus on ABC...

- We explore the use of these approximate methods to conduct **forecasting** in intractable and ***misspecified*** SSMs
- with the aim of producing accurate, fit for purpose predictions
- We are explicitly looking at the **use of ABC**
- The **aim of ABC**

We focus on ABC...

- We explore the use of these approximate methods to conduct **forecasting** in intractable and ***misspecified*** SSMs
- with the aim of producing accurate, fit for purpose predictions
- We are explicitly looking at the **use of ABC**
- The **aim of ABC**
    - is to produce draws from an **approximation** to $p(\theta|y_{1:T})$

We focus on ABC...

- We explore the use of these approximate methods to conduct **forecasting** in intractable and ***misspecified*** SSMs
- with the aim of producing accurate, fit for purpose predictions
- We are explicitly looking at the **use of ABC**
- The **aim of ABC**
  - is to produce draws from an **approximation** to $p(\theta|y_{1:T})$
  - in the case where DGP cannot be evaluated

We focus on ABC...

- We explore the use of these approximate methods to conduct **forecasting** in intractable and ***misspecified*** SSMs
- with the aim of producing accurate, fit for purpose predictions
- We are explicitly looking at the **use of ABC**
- The **aim of ABC**
  - is to produce draws from an **approximation** to $p(\theta|y_{1:T})$
  - in the case where DGP cannot be evaluated
  - But *can* be **simulated** from

ABC Algorithm (adapted for SSM)

ABC Algorithm (adapted for SSM)

1. Simulate $i = 1, 2, ..., N$, $i.i.d.$ draws of $\theta^{(i)}$ from $p(\theta)$

ABC Algorithm (adapted for SSM)

1. Simulate $i = 1, 2, ..., N$, $i.i.d.$ draws of $\theta^{(i)}$ from $p(\theta)$
2. Simulate

ABC Algorithm (adapted for SSM)

1. Simulate $i = 1, 2, ..., N$, *i.i.d.* draws of $\theta^{(i)}$ from $p(\theta)$
2. Simulate
   - **pseudo-states** $x^s_{0:T}(\theta^i)$, $i = 1, 2, ..., N$ from $p(x^s_0|\theta^i)$ and $p(x^s_t|x^s_{t-1}, \theta^i)$, for $t = 1, 2, ..., T$.

ABC Algorithm (adapted for SSM)

1. Simulate $i = 1, 2, ..., N$, *i.i.d.* draws of $\theta^{(i)}$ from $p(\theta)$
2. Simulate
   - **pseudo-states** $x^s_{0:T}(\theta^i), i = 1, 2, ..., N$ from $p(x^s_0|\theta^i)$ and $p(x^s_t|x^s_{t-1}, \theta^i)$, for $t = 1, 2, ..., T$.
   - **pseudo-data** $y^s_{1:T}(\theta^i), i = 1, 2, ..., N$ from $p(y^s_t|x^s_t, \theta^i)$, for $t = 1, 2, ..., T$.

ABC Algorithm (adapted for SSM)

1. Simulate $i = 1, 2, ..., N$, $i.i.d.$ draws of $\theta^{(i)}$ from $p(\theta)$
2. Simulate
   - **pseudo-states** $x^s_{0:T}(\theta^i)$, $i = 1, 2, ..., N$ from $p(x^s_0|\theta^i)$ and $p(x^s_t|x^s_{t-1}, \theta^i)$, for $t = 1, 2, ..., T$.
   - **pseudo-data** $y^s_{1:T}(\theta^i)$, $i = 1, 2, ..., N$ from $p(y^s_t|x^s_t, \theta^i)$, for $t = 1, 2, ..., T$.
3. Select $\theta^{(i)}$ such that:

$$d\{\eta(y_{1:T}), \eta\left(y^s_{1:T}(\theta^i)\right)\} \leq \varepsilon,$$

ABC Algorithm (adapted for SSM)

1. Simulate $i = 1, 2, ..., N$, *i.i.d.* draws of $\theta^{(i)}$ from $p(\theta)$
2. Simulate
   - **pseudo-states** $x^s_{0:T}(\theta^i)$, $i = 1, 2, ..., N$ from $p(x^s_0|\theta^i)$ and $p(x^s_t|x^s_{t-1}, \theta^i)$, for $t = 1, 2, ..., T$.
   - **pseudo-data** $y^s_{1:T}(\theta^i)$, $i = 1, 2, ..., N$ from $p(y^s_t|x^s_t, \theta^i)$, for $t = 1, 2, ..., T$.
3. Select $\theta^{(i)}$ such that:

$$d\{\eta(y_{1:T}), \eta\left(y^s_{1:T}(\theta^i)\right)\} \leq \varepsilon,$$

   - $\eta(.)$ is a (vector) **summary statistic**

ABC Algorithm (adapted for SSM)

1. Simulate $i = 1, 2, ..., N$, *i.i.d.* draws of $\theta^{(i)}$ from $p(\theta)$
2. Simulate
   - **pseudo-states** $x^s_{0:T}(\theta^i)$, $i = 1, 2, ..., N$ from $p(x^s_0|\theta^i)$ and $p(x^s_t|x^s_{t-1}, \theta^i)$, for $t = 1, 2, ..., T$.
   - **pseudo-data** $y^s_{1:T}(\theta^i)$, $i = 1, 2, ..., N$ from $p(y^s_t|x^s_t, \theta^i)$, for $t = 1, 2, ..., T$.
3. Select $\theta^{(i)}$ such that:

$$d\{\eta(y_{1:T}), \eta\left(y^s_{1:T}(\theta^i)\right)\} \leq \varepsilon,$$

   - $\eta(.)$ is a (vector) **summary statistic**
   - $d\{.\}$ is a **distance criterion**

ABC Algorithm (adapted for SSM)

1. Simulate $i = 1, 2, ..., N$, $i.i.d.$ draws of $\theta^{(i)}$ from $p(\theta)$
2. Simulate
   - **pseudo-states** $x^s_{0:T}(\theta^i)$, $i = 1, 2, ..., N$ from $p(x^s_0|\theta^i)$ and $p(x^s_t|x^s_{t-1}, \theta^i)$, for $t = 1, 2, ..., T$.
   - **pseudo-data** $y^s_{1:T}(\theta^i)$, $i = 1, 2, ..., N$ from $p(y^s_t|x^s_t, \theta^i)$, for $t = 1, 2, ..., T$.
3. Select $\theta^{(i)}$ such that:

$$d\{\eta(y_{1:T}), \eta\left(y^s_{1:T}(\theta^i)\right)\} \leq \varepsilon,$$

   - $\eta(.)$ is a (vector) **summary statistic**
   - $d\{.\}$ is a **distance criterion**
   - the **tolerance** ε is arbitrarily small

ABC Algorithm (adapted for SSM)

1. Simulate $i = 1, 2, ..., N$, *i.i.d.* draws of $\theta^{(i)}$ from $p(\theta)$
2. Simulate
   - **pseudo-states** $x^s_{0:T}(\theta^i)$, $i = 1, 2, ..., N$ from $p(x^s_0|\theta^i)$ and $p(x^s_t|x^s_{t-1}, \theta^i)$, for $t = 1, 2, ..., T$.
   - **pseudo-data** $y^s_{1:T}(\theta^i)$, $i = 1, 2, ..., N$ from $p(y^s_t|x^s_t, \theta^i)$, for $t = 1, 2, ..., T$.
3. Select $\theta^{(i)}$ such that:

$$d\{\eta(y_{1:T}), \eta\left(y^s_{1:T}(\theta^i)\right)\} \leq \varepsilon,$$

   - $\eta(.)$ is a (vector) **summary statistic**
   - $d\{.\}$ is a **distance criterion**
   - the **tolerance** ε is arbitrarily small
4. Selected draws $\Rightarrow$ draws from $p_\varepsilon(\theta|\eta(y_{1:T}))$

Existing ABC Literature

Existing ABC Literature

- Under **correct model specification,**

Existing ABC Literature

- Under **correct model specification**,

Existing ABC Literature

- Under **correct model specification**,
  - ABC posterior - Bayesian consistent for the true parameter, asymptotically normal
    [Frazier et al., 2018] - 'Asymptotic properties of approximate Bayesian computation'

Existing ABC Literature

- Under **correct model specification**,
  - ABC posterior - Bayesian consistent for the true parameter, asymptotically normal
    [Frazier et al., 2018] - 'Asymptotic properties of approximate Bayesian computation'
  - also in an explicitly SSM setting
    [Martin et al., 2019] - 'Auxiliary likelihood-based approximate Bayesian computation in state space models'

Existing ABC Literature

- Under **correct model specification**,
  - ABC posterior - Bayesian consistent for the true parameter, asymptotically normal
    [Frazier et al., 2018] - 'Asymptotic properties of approximate Bayesian computation'
  - also in an explicitly SSM setting
    [Martin et al., 2019] - 'Auxiliary likelihood-based approximate Bayesian computation in state space models'
  - ABC-based forecasting
    [Frazier et al., 2019] - 'Approximate Bayesian forecasting'

Existing ABC Literature

- Under **correct model specification**,
  - ABC posterior - Bayesian consistent for the true parameter, asymptotically normal
    [Frazier et al., 2018] - 'Asymptotic properties of approximate Bayesian computation'
  - also in an explicitly SSM setting
    [Martin et al., 2019] - 'Auxiliary likelihood-based approximate Bayesian computation in state space models'
  - ABC-based forecasting
    [Frazier et al., 2019] - 'Approximate Bayesian forecasting'
- Under **model misspecification**

Existing ABC Literature

- Under **correct model specification**,
  - ABC posterior - Bayesian consistent for the true parameter, asymptotically normal
    [Frazier et al., 2018] - 'Asymptotic properties of approximate Bayesian computation'
  - also in an explicitly SSM setting
    [Martin et al., 2019] - 'Auxiliary likelihood-based approximate Bayesian computation in state space models'
  - ABC-based forecasting
    [Frazier et al., 2019] - 'Approximate Bayesian forecasting'
- Under **model misspecification**
  - **[Frazier et al., 2020] - 'Model misspecification in approximate Bayesian computation: consequences and diagnostics'**

What is left unexplored?

What is left unexplored?

- **ABC-based forecasting** – with *misspecified* SSMs

What is left unexplored?

- **ABC-based forecasting** – with ***misspecified*** SSMs
- If we ***do not assume*** the correct model specification $\Rightarrow$

What is left unexplored?

- **ABC-based forecasting** – with ***misspecified*** SSMs
- If we ***do not assume*** the correct model specification $\Rightarrow$
- the whole focus of the Bayesian forecasting exercise **needs to change**.

What is left unexplored?

- **ABC-based forecasting** – with ***misspecified*** SSMs
- If we ***do not assume*** the correct model specification ⇒
- the whole focus of the Bayesian forecasting exercise **needs to change**.
  - argument put forward in recent forecasting works by
    - **[Loaiza-Maya et al., 2021a] - 'Focused Bayesian prediction'**
    - **[Frazier et al., 2021] - 'Loss-based variational Bayes prediction'**

Loss-based Bayesian Prediction

Loss-based Bayesian Prediction

- **Essence of the idea:**

Loss-based Bayesian Prediction

- **Essence of the idea:**
- In the spirit of the various generalized Bayesian *inferential* methods,
  - **[Bissiri et al., 2016]; [Giummolè et al., 2019]; [Knoblauch et al., 2019]; [Pacchiardi and Dutta, 2021]; [Loaiza-Maya et al., 2021a] and [Frazier et al., 2021]**

Loss-based Bayesian Prediction

- **Essence of the idea:**
- In the spirit of the various generalized Bayesian *inferential* methods,
  - **[Bissiri et al., 2016]; [Giummolè et al., 2019]; [Knoblauch et al., 2019]; [Pacchiardi and Dutta, 2021]; [Loaiza-Maya et al., 2021a] and [Frazier et al., 2021]**
- replace the **likelihood function (log-score loss)** in the conventional Bayesian update

Loss-based Bayesian Prediction

- **Essence of the idea:**
- In the spirit of the various generalized Bayesian *inferential* methods,
  - **[Bissiri et al., 2016]; [Giummolè et al., 2019]; [Knoblauch et al., 2019]; [Pacchiardi and Dutta, 2021]; [Loaiza-Maya et al., 2021a] and [Frazier et al., 2021]**
- replace the **likelihood function (log-score loss)** in the conventional Bayesian update
- by the *particular* **predictive loss** that matters for the *particular* forecasting problem being tackled

Loss-based Bayesian Prediction in SSMs

- Assume a class of plausible **predictive SSMs** for Y_{T+1}, conditioned on the information $\mathcal{F}_T$:

$$\mathcal{P}^{(T)} := \{P_\theta^{(T)}, \theta \in \Theta\}$$

Loss-based Bayesian Prediction in SSMs

- Assume a class of plausible **predictive SSMs** for Y_{T+1}, conditioned on the information $\mathcal{F}_T$:

$$\mathcal{P}^{(T)} := \{P_\theta^{(T)}, \theta \in \Theta\}$$

- Assume that the model is **misspecified**

Loss-based Bayesian Prediction in SSMs

- Assume a class of plausible **predictive SSMs** for Y_{T+1}, conditioned on the information $\mathcal{F}_T$:

$$\mathcal{P}^{(T)} := \{P_{\theta}^{(T)}, \theta \in \Theta\}$$

- Assume that the model is **misspecified**
  - Likelihood based inference: $P_0 \notin \mathcal{P}^{(T)}$

Loss-based Bayesian Prediction in SSMs

- Assume a class of plausible **predictive SSMs** for Y_{T+1}, conditioned on the information $\mathcal{F}_T$:

$$\mathcal{P}^{(T)} := \{P_\theta^{(T)}, \theta \in \Theta\}$$

- Assume that the model is **misspecified**
  - Likelihood based inference: $P_0 \notin \mathcal{P}^{(T)}$
  - But... in ABC
    - there does not exist any $\theta_0 \in \Theta$ satisfying $b_0 = b(\theta_0)$
    - where it is assumed that the summary statistics converge to some fixed value, namely b_0 under P_0 and $b(\theta)$ under $P_\theta^{(T)}$ **[Frazier et al., 2020]**

Loss-based Bayesian Prediction in SSMs

- Construct the **loss-based posterior/Gibbs posterior** $p_L(\theta|y_{1:T})$

Loss-based Bayesian Prediction in SSMs

- Construct the **loss-based posterior/Gibbs posterior** $p_L(\theta|y_{1:T})$
  - via some (positively-oriented) **scoring rule**

$$S_T(\theta) := \sum_{t=0}^{T-1} s(P_\theta^{(t)}, y_{t+1})$$

Loss-based Bayesian Prediction in SSMs

- Construct the **loss-based posterior/Gibbs posterior** $p_L(\theta|y_{1:T})$
  - via some (positively-oriented) **scoring rule**

$$S_T(\theta) := \sum_{t=0}^{T-1} s(P_\theta^{(t)}, y_{t+1})$$

- **Loss-based predictive**

$$\begin{aligned} & p_L(y_{T+1}|y_{1:T}) \\ = \quad & \int_{x_{T+1}} \int_{x_{0:T}} \int_\theta p(y_{T+1}|x_{T+1}, \theta, y_{1:T}) p(x_{T+1}|x_T, \theta) \\ & \times p(x_{0:T}|y_{1:T}, \theta) p_L(\theta|y_{1:T}) d\theta dx_{0:T} dx_{T+1} \end{aligned}$$

Immediate Problems...

Immediate Problems...

1. Consider the **loss-based predictive** $p_L(y_{T+1}|y_{1:T})$ which is constructed based on $p_L(\theta|y_{1:T})$ via

$$S_T(\theta) := \sum_{t=0}^{T-1} s(\underbrace{p(y_{t+1}|y_{1:T},\theta)}_{\text{with marginal of } x_{0:T}}, y_{t+1})$$

Immediate Problems...

1. Consider the **loss-based predictive** $p_L(y_{T+1}|y_{1:T})$ which is constructed based on $p_L(\theta|y_{1:T})$ via

$$S_T(\theta) := \sum_{t=0}^{T-1} s(\underbrace{p(y_{t+1}|y_{1:T},\theta)}_{\text{with marginal of } x_{0:T}}, y_{t+1})$$

   - Intractable SSM model $\Rightarrow$ $p_L(\theta|y_{1:T})$ out of reach

Immediate Problems...

1. Consider the **loss-based predictive** $p_L(y_{T+1}|y_{1:T})$ which is constructed based on $p_L(\theta|y_{1:T})$ via

$$S_T(\theta) := \sum_{t=0}^{T-1} s\Big(\underbrace{p(y_{t+1}|y_{1:T},\theta)}_{\text{with marginal of } x_{0:T}}, y_{t+1}\Big)$$

   - Intractable SSM model $\Rightarrow p_L(\theta|y_{1:T})$ out of reach

2. In a **likelihood-based** setting we deal with the **joint (augmented)** posterior:

$$p(x_{0:T},\theta|y_{1:T}) = \underbrace{p(x_{0:T}|y_{1:T},\theta)}\,\underbrace{p\,(\theta|y_{1:T})}$$

(typically) via MCMC

Immediate Problems...

- In the **loss-based** setting the decomposition:

$$\underbrace{p(x_{0:T}|y_{1:T},\theta)}\,\underbrace{p_L\left(\theta|y_{1:T}\right)}$$

Immediate Problems...

- In the **loss-based** setting the decomposition:

$$\underbrace{p(x_{0:T}|y_{1:T},\theta)}\underbrace{p_L(\theta|y_{1:T})}$$

- $\Rightarrow p(x_{0:T}|y_{1:T},\theta)$ and $p_L(\theta|y_{1:T})$ need to be tackled 'on their own'

Our solutions..

Our solutions..

1. Approximate $p_L(\theta|y_{1:T})$ using **ABC**

Our solutions..

1. Approximate $p_L\left(\theta|y_{1:T}\right)$ using **ABC**
   - $\Rightarrow$ Enables marginalization w.r.t. $x_{0:T}$

Our solutions..

1. Approximate $p_L\left(\theta | y_{1:T}\right)$ using **ABC**
   - $\Rightarrow$ Enables marginalization w.r.t. $x_{0:T}$
   - **NOTE:** this is useful ***even when the DGP itself is tractable***

Our solutions..

1. Approximate $p_L\left(\theta|y_{1:T}\right)$ using **ABC**
   - $\Rightarrow$ Enables marginalization w.r.t. $x_{0:T}$
   - **NOTE:** this is useful ***even when the DGP itself is tractable***

2. Recognize that $p(x_{0:T}|y_{1:T},\theta)$ is not required for **prediction** in an **SSM**

Our solutions..

1. Approximate $p_L(\theta|y_{1:T})$ using **ABC**
   - $\Rightarrow$ Enables marginalization w.r.t. $x_{0:T}$
   - **NOTE:** this is useful ***even when the DGP itself is tractable***

2. Recognize that $p(x_{0:T}|y_{1:T},\theta)$ is not required for **prediction** in an **SSM**
   - Only need to access $p(x_T|y_{1:T},\theta)$

Our solutions..

1. Approximate $p_L\left(\theta|y_{1:T}\right)$ using **ABC**
    - $\Rightarrow$ Enables marginalization w.r.t. $x_{0:T}$
    - **NOTE:** this is useful ***even when the DGP itself is tractable***

2. Recognize that $p(x_{0:T}|y_{1:T}, \theta)$ is not required for **prediction** in an **SSM**
    - Only need to access $p(x_T|y_{1:T}, \theta)$
    - Can be achieved **exactly** via particle filtering

Our solutions..

1. Approximate $p_L(\theta|y_{1:T})$ using **ABC**
   - $\Rightarrow$ Enables marginalization w.r.t. $x_{0:T}$
   - **NOTE:** this is useful ***even when the DGP itself is tractable***
2. Recognize that $p(x_{0:T}|y_{1:T},\theta)$ is not required for **prediction** in an **SSM**
   - Only need to access $p(x_T|y_{1:T},\theta)$
   - Can be achieved **exactly** via particle filtering
   - Conditional on any draw of θ

How Do We Choose the Summary Statistics?

How Do We Choose the Summary Statistics?

- In the spirit of ***'auxiliary model'***- based ABC :
  - **[Drovandi et al., 2011], [Drovandi et al., 2015] and [Martin et al., 2019]**

How Do We Choose the Summary Statistics?

- In the spirit of *'auxiliary model'*- based ABC :
  - **[Drovandi et al., 2011], [Drovandi et al., 2015] and [Martin et al., 2019]**
- We choose an **auxiliary** model

How Do We Choose the Summary Statistics?

- In the spirit of ***'auxiliary model'***- based ABC :
  - **[Drovandi et al., 2011], [Drovandi et al., 2015] and [Martin et al., 2019]**
- We choose an **auxiliary** model
  - that is a 'reasonable' approximation to the assumed **SSM** &

How Do We Choose the Summary Statistics?

- In the spirit of ***'auxiliary model'***- based ABC :
  - **[Drovandi et al., 2011], [Drovandi et al., 2015] and [Martin et al., 2019]**
- We choose an **auxiliary** model
  - that is a 'reasonable' approximation to the assumed **SSM** &
  - admit a **closed-form predictive** (with parameter vector β)

How Do We Choose the Summary Statistics?

- In the spirit of ***'auxiliary model'***- based ABC :
  - **[Drovandi et al., 2011], [Drovandi et al., 2015] and [Martin et al., 2019]**
- We choose an **auxiliary** model
  - that is a 'reasonable' approximation to the assumed **SSM** &
  - admit a **closed-form predictive** (with parameter vector β)
- Define the score-based criterion as:

$$S_T(\beta) := \sum_{t=0}^{T-1} s(P_\beta^{(t)}, y_{t+1})$$

How Do We Choose the Summary Statistics?

- ***Summary statistics*** : average of the first-derivative of $S_T(\beta)$ using a scoring rule s_j

How Do We Choose the Summary Statistics?

- ***Summary statistics*** : average of the first-derivative of $S_T(\beta)$ using a scoring rule s_j
  - evaluated at $\hat{\beta}_j(y_{1:T})$: optimizer of $S_T(\beta)$ using s_j

How Do We Choose the Summary Statistics?

- ***Summary statistics*** : average of the first-derivative of $S_T(\beta)$ using a scoring rule s_j
  - evaluated at $\hat{\beta}_j(y_{1:T})$: optimizer of $S_T(\beta)$ using s_j
- That is,

$$\eta_j(y^s_{1:T}) = T^{-1} \frac{\partial \sum_{t=0}^{T-1} s_j(P_\beta^{(t)}, y^s_{t+1}(\theta^i))}{\partial \beta}\Bigg|_{\beta=\hat{\beta}_j(y_{1:T})}$$

and:

$$\eta_j(y_{1:T}) = T^{-1} \frac{\partial \sum_{t=0}^{T-1} s_j(P_\beta^{(t)}, y_{t+1})}{\partial \beta}\Bigg|_{\beta=\hat{\beta}_j(y_{1:T})} = 0$$

Distance Criterion

- *ABC distance* : Mahalanobis distance

$$d\{\eta\left(y_{1:T}^{s}\right),\eta\left(y_{1:T}\right)\} = \sqrt{\left[\bar{S}_j\left\{y_{1:T}^{s}(\theta^i);\hat{\beta}_j(y_{1:T})\right\}\right]'\hat{\Sigma}\left[\bar{S}_j\left\{y_{1:T}^{s}(\theta^i);\hat{\beta}_j(y_{1:T})\right\}\right]},$$

  - where $\hat{\Sigma}$ is the inverse of the (estimated) covariance matrix of $\eta_j(y_{1:T}^{s})$ across draws and
  - $\bar{S}_j\left\{y_{1:T}^{s}(\theta^i);\hat{\beta}_j(y_{1:T})\right\} = T^{-1}\left.\frac{\partial\sum_{t=0}^{T-1}s_j(P_{\beta}^{(t)},y_{t+1}^{s}(\theta^i))}{\partial\beta}\right|_{\beta=\hat{\beta}_j(y_{1:T})}$

Loss-based ABF

- The accepted draws of θ produced by the ABC algorithm $\Rightarrow$

Loss-based ABF

- The accepted draws of θ produced by the ABC algorithm $\Rightarrow$
- are the draws from **loss-based ABC posterior,** $p_{L,\varepsilon}(\theta|\eta(y_{1:T}))$

Loss-based ABF

- The accepted draws of θ produced by the ABC algorithm $\Rightarrow$
- are the draws from **loss-based ABC posterior,** $p_{L,\varepsilon}(\theta|\eta(y_{1:T}))$
- **Loss-based ABC predictive** :

$$
\begin{aligned}
& g_L(y_{T+1}|y_{1:T}) \\
= \quad & \int_{x_{T+1}}\int_{x_T}\int_{\theta} p(y_{T+1}|x_{T+1},\theta,y_{1:T})p(x_{T+1}|x_T,\theta)p(x_T|\theta,y_{1:T}) \\
\times \quad & p_{L,\varepsilon}(\theta|\eta(y_{1:T}))d\theta dx_T dx_{T+1}.
\end{aligned}
$$

Loss-based ABF

- The accepted draws of θ produced by the ABC algorithm $\Rightarrow$
- are the draws from **loss-based ABC posterior,** $p_{L,\varepsilon}(\theta|\eta(y_{1:T}))$
- **Loss-based ABC predictive** :

$$\begin{aligned} & g_L(y_{T+1}|y_{1:T}) \\ = & \int_{x_{T+1}} \int_{x_T} \int_{\theta} p(y_{T+1}|x_{T+1}, \theta, y_{1:T}) p(x_{T+1}|x_T, \theta) p(x_T|\theta, y_{1:T}) \\ \times & \; p_{L,\varepsilon}(\theta|\eta(y_{1:T})) d\theta dx_T dx_{T+1}. \end{aligned}$$

- $\Rightarrow$ loss-based ABC forecasting (**loss-based ABF**)

Numerical Illustration: Simulation design

Numerical Illustration: Simulation design

- **Predictive class**, $P_\theta^{(t)}$ is a SSM:

Numerical Illustration: Simulation design

- **Predictive class**, $P_{\theta}^{(t)}$ is a SSM:
  - **SV model** for a continuous financial return, y_t,

Numerical Illustration: Simulation design

- **Predictive class**, $P_\theta^{(t)}$ is a SSM:
  - **SV model** for a continuous financial return, y_t,

Numerical Illustration: Simulation design

- **Predictive class**, $P_\theta^{(t)}$ is a SSM:
  - **SV model** for a continuous financial return, y_t,

$$y_t = \mu + e^{\alpha_t/2} e_t \quad ; \quad e_t \sim N(0,1)$$

Numerical Illustration: Simulation design

- **Predictive class**, $P_{\theta}^{(t)}$ is a SSM:
  - **SV model** for a continuous financial return, y_t,

$$y_t = \mu + e^{\alpha_t/2} e_t \quad ; \quad e_t \sim N(0,1)$$

$$\alpha_t = \bar{h}_\alpha + \phi(\alpha_{t-1} - \bar{h}_\alpha) + w_t \quad ; \quad w_t \sim N(0, \sigma_\alpha^2)$$

Numerical Illustration: Simulation design

- **Predictive class**, $P_{\theta}^{(t)}$ is a SSM:
  - **SV model** for a continuous financial return, y_t,

$$y_t = \mu + e^{\alpha_t/2} e_t \quad ; \quad e_t \sim N(0,1)$$

$$\alpha_t = \bar{h}_\alpha + \phi(\alpha_{t-1} - \bar{h}_\alpha) + w_t \quad ; \quad w_t \sim N(0, \sigma_\alpha^2)$$

$$\alpha_0 \sim N\left(\bar{h}_\alpha, \frac{\sigma_\alpha^2}{1-\phi^2}\right),$$

Numerical Illustration: Simulation design

- **Predictive class**, $P_\theta^{(t)}$ is a SSM:
  - **SV model** for a continuous financial return, y_t,

$$y_t = \mu + e^{\alpha_t/2} e_t \quad ; \quad e_t \sim N(0,1)$$

$$\alpha_t = \bar{h}_\alpha + \phi(\alpha_{t-1} - \bar{h}_\alpha) + w_t \quad ; \quad w_t \sim N(0, \sigma_\alpha^2)$$

$$\alpha_0 \sim N\left(\bar{h}_\alpha, \frac{\sigma_\alpha^2}{1-\phi^2}\right),$$

- $\theta = (\phi, \sigma_\alpha^2, \mu, \bar{h}_\alpha)'$

Numerical Illustration: Simulation design

- **True DGP**

Numerical Illustration: Simulation design

- **True DGP**
  1. A model that matches the assumed SV model

Numerical Illustration: Simulation design

- **True DGP**
  1. A model that matches the assumed SV model
     - $\Rightarrow$ **Correct model specification**

Numerical Illustration: Simulation design

- **True DGP**
  1. A model that matches the assumed SV model
     - $\Rightarrow$ **Correct model specification**
  2. SV model that better replicates the stylized features of financial returns data, as used by **[Loaiza-Maya et al., 2021a]** :

$$z_t = exp(0.5h_t)\epsilon_t$$

$$h_t = \bar{h} + a(h_{t-1} - \bar{h}) + \sigma_h \eta_t$$

$$y_t = G^{-1}(F_z(z_t))$$

Numerical Illustration: Simulation design

- **True DGP**
  1. A model that matches the assumed SV model
     - $\Rightarrow$ **Correct model specification**
  2. SV model that better replicates the stylized features of financial returns data, as used by **[Loaiza-Maya et al., 2021a]** :

$$z_t = exp(0.5h_t)\epsilon_t$$
$$h_t = \bar{h} + a(h_{t-1} - \bar{h}) + \sigma_h \eta_t$$
$$y_t = G^{-1}(F_z(z_t))$$

     - $\Rightarrow$ Implied copula of a **stochastic volatility** model combined with a **skewed normal marginal,** $g(y_t)$ (imposed via G^{-1})

Numerical Illustration: Simulation design

- **True DGP**
  1. A model that matches the assumed SV model
     - $\Rightarrow$ **Correct model specification**
  2. SV model that better replicates the stylized features of financial returns data, as used by **[Loaiza-Maya et al., 2021a]** :

$$z_t = exp(0.5h_t)\epsilon_t$$
$$h_t = \bar{h} + a(h_{t-1} - \bar{h}) + \sigma_h \eta_t$$
$$y_t = G^{-1}(F_z(z_t))$$

     - $\Rightarrow$ Implied copula of a **stochastic volatility** model combined with a **skewed normal marginal,** $g(y_t)$ (imposed via G^{-1})
     - $\Rightarrow$ **Model misspecification**

Numerical Illustration: Simulation design

- **True DGP**
  1. A model that matches the assumed SV model
     - $\Rightarrow$ **Correct model specification**
  2. SV model that better replicates the stylized features of financial returns data, as used by [Loaiza-Maya et al., 2021a] :

$$z_t = exp(0.5h_t)\epsilon_t$$
$$h_t = \bar{h} + a(h_{t-1} - \bar{h}) + \sigma_h \eta_t$$
$$y_t = G^{-1}(F_z(z_t))$$

     - $\Rightarrow$ Implied copula of a **stochastic volatility** model combined with a **skewed normal marginal,** $g(y_t)$ (imposed via G^{-1})
     - $\Rightarrow$ **Model misspecification**
- Predicting **extreme returns** accurately is important

Numerical Illustration: Simulation design

- **True DGP**
  1. A model that matches the assumed SV model
     - $\Rightarrow$ **Correct model specification**
  2. SV model that better replicates the stylized features of financial returns data, as used by **[Loaiza-Maya et al., 2021a]** :

$$z_t = exp(0.5h_t)\epsilon_t$$
$$h_t = \bar{h} + a(h_{t-1} - \bar{h}) + \sigma_h \eta_t$$
$$y_t = G^{-1}(F_z(z_t))$$

     - $\Rightarrow$ Implied copula of a **stochastic volatility** model combined with a **skewed normal marginal,** $g(y_t)$ (imposed via G^{-1})
     - $\Rightarrow$ **Model misspecification**
- Predicting **extreme returns** accurately is important
- Will **focus** on that goal $\Rightarrow$ use an appropriate s_j in the up-date

Numerical Illustration: Simulation design

- Three **auxiliary models** are used:
  1. a Gaussian ARCH(1) model
  2. a Gaussian GARCH(1,1) model
  3. a GARCH(1,1) model with Student-t errors

Numerical Illustration: Simulation design

- Three **auxiliary models** are used:
  1. a Gaussian ARCH(1) model
  2. a Gaussian GARCH(1,1) model
  3. a GARCH(1,1) model with Student-t errors

- Results for three types of **scores** reproduced here:

Numerical Illustration: Simulation design

- Three **auxiliary models** are used:
  1. a Gaussian ARCH(1) model
  2. a Gaussian GARCH(1,1) model
  3. a GARCH(1,1) model with Student-t errors

- Results for three types of **scores** reproduced here:
  1. Log score (LS)

Numerical Illustration: Simulation design

- Three **auxiliary models** are used:
  1. a Gaussian ARCH(1) model
  2. a Gaussian GARCH(1,1) model
  3. a GARCH(1,1) model with Student-t errors

- Results for three types of **scores** reproduced here:
  1. Log score (LS)
  2. **Censored** log score (CLS) (rewards **predictive accuracy in a tail**)

Numerical Illustration: Simulation design

- Three **auxiliary models** are used:
  1. a Gaussian ARCH(1) model
  2. a Gaussian GARCH(1,1) model
  3. a GARCH(1,1) model with Student-t errors

- Results for three types of **scores** reproduced here:
  1. Log score (LS)
  2. **Censored** log score (CLS) (rewards **predictive accuracy in a tail**)
  3. Continuously ranked probability score (CRPS) (rewards the assignment of high predictive mass near to the realized value)

Numerical Illustration: Simulation design

- Three **auxiliary models** are used:
  1. a Gaussian ARCH(1) model
  2. a Gaussian GARCH(1,1) model
  3. a GARCH(1,1) model with Student-t errors

- Results for three types of **scores** reproduced here:
  1. Log score (LS)
  2. **Censored** log score (CLS) (rewards **predictive accuracy in a tail**)
  3. Continuously ranked probability score (CRPS) (rewards the assignment of high predictive mass near to the realized value)
  - [Gneiting and Raftery, 2007]

Numerical Illustration: Simulation design

- Three **auxiliary models** are used:
  1. a Gaussian ARCH(1) model
  2. a Gaussian GARCH(1,1) model
  3. a GARCH(1,1) model with Student-t errors

- Results for three types of **scores** reproduced here:
  1. Log score (LS)
  2. **Censored** log score (CLS) (rewards **predictive accuracy in a tail**)
  3. Continuously ranked probability score (CRPS) (rewards the assignment of high predictive mass near to the realized value)
     - [Gneiting and Raftery, 2007]

- Will discuss the results based on **Gaussian GARCH(1,1)** model under **model misspecification**

Preliminary Results: Model misspecification

Preliminary Results: Model misspecification

Average out-of-sample score

ABC Score	LS	$CS_{<10\%}$	$CS_{>80\%}$	$CS_{>90\%}$	CRPS
LS	*-1.343*	-0.359	-0.490	-0.298	-0.533
$CS_{<10\%}$	-1.412	*-0.357*	-0.512	-0.304	-0.562
$CS_{>80\%}$	-2.092	-0.812	**-0.468**	-0.282	-0.608
$CS_{>90\%}$	-2.426	-0.896	-0.472	**-0.278**	-0.651
CRPS	-1.337	-0.363	-0.488	-0.299	**-0.530**
Exact Bayes	-1.334	**-0.362**	**-0.488**	**-0.300**	**-0.530**
True	-1.183	-0.328	-0.407	-0.230	-0.514

Preliminary Results: Model misspecification

Average out-of-sample score

ABC Score	LS	$CS_{<10\%}$	$CS_{>80\%}$	$CS_{>90\%}$	CRPS
LS	*-1.343*	-0.359	-0.490	-0.298	-0.533
$CS_{<10\%}$	-1.412	*-0.357*	-0.512	-0.304	-0.562
$CS_{>80\%}$	-2.092	-0.812	**-0.468**	-0.282	-0.608
$CS_{>90\%}$	-2.426	-0.896	-0.472	**-0.278**	-0.651
CRPS	-1.337	-0.363	-0.488	-0.299	**-0.530**
Exact Bayes	-1.334	**-0.362**	**-0.488**	**-0.300**	**-0.530**
True	-1.183	-0.328	-0.407	-0.230	-0.514

- Positively-oriented scores ⇒ large (**in bold**) is good

Preliminary Results: Model misspecification

Average out-of-sample score

ABC Score	LS	$CS_{<10\%}$	$CS_{>80\%}$	$CS_{>90\%}$	CRPS
LS	*-1.343*	-0.359	-0.490	-0.298	-0.533
$CS_{<10\%}$	-1.412	*-0.357*	-0.512	-0.304	-0.562
$CS_{>80\%}$	-2.092	-0.812	**-0.468**	-0.282	-0.608
$CS_{>90\%}$	-2.426	-0.896	-0.472	**-0.278**	-0.651
CRPS	-1.337	-0.363	-0.488	-0.299	**-0.530**
Exact Bayes	-1.334	**-0.362**	**-0.488**	**-0.300**	**-0.530**
True	-1.183	-0.328	-0.407	-0.230	-0.514

- Positively-oriented scores $\Rightarrow$ large (**in bold**) is good
- Looking for **bold** values on the diagonal

Preliminary Results: Model misspecification

Average out-of-sample score

ABC Score	LS	$CS_{<10\%}$	$CS_{>80\%}$	$CS_{>90\%}$	CRPS
LS	*-1.343*	-0.359	-0.490	-0.298	-0.533
$CS_{<10\%}$	-1.412	*-0.357*	-0.512	-0.304	-0.562
$CS_{>80\%}$	-2.092	-0.812	**-0.468**	-0.282	-0.608
$CS_{>90\%}$	-2.426	-0.896	-0.472	**-0.278**	-0.651
CRPS	-1.337	-0.363	-0.488	-0.299	**-0.530**
Exact Bayes	-1.334	**-0.362**	**-0.488**	**-0.300**	**-0.530**
True	-1.183	-0.328	-0.407	-0.230	-0.514

- Positively-oriented scores ⇒ large (**in bold**) is good
- Looking for **bold** values on the diagonal
- **loss-based ABF** > **misspecified exact Bayes**

Preliminary Results: Model misspecification

- **Coherent predictions** are in evidence

Preliminary Results: Model misspecification

- **Coherent predictions** are in evidence

Preliminary Results: Model misspecification

- **Coherent predictions** are in evidence
  - a loss-based ABC predictive constructed via a particular scoring rule
  - performs the best out-of-sample according to that same score
  - when compared with a loss-based ABC predictive constructed via some different scoring rule

Theoretical Validation

Theoretical Validation

- Given that we are using ABC in a misspecified setting

Theoretical Validation

- Given that we are using ABC in a misspecified setting
- **ABC inference** is adversely affected by misspecification **[Frazier et al., 2020]**

Theoretical Validation

- Given that we are using ABC in a misspecified setting
- **ABC inference** is adversely affected by misspecification **[Frazier et al., 2020]**
  - ABC posterior does concentrate all mass on an appropriately defined pseudo-true value $\theta^\star$

$$\theta^\star = \arg\min_{\theta \in \Theta} d\{b_0, b(\theta)\},$$

Theoretical Validation

- Given that we are using ABC in a misspecified setting
- **ABC inference** is adversely affected by misspecification **[Frazier et al., 2020]**
  - ABC posterior does concentrate all mass on an appropriately defined pseudo-true value $\theta^\star$

$$\theta^\star = \arg\min_{\theta \in \Theta} d\{b_0, b(\theta)\},$$

  - But, the posterior does not concentrate in a Gaussian manner

Theoretical Validation

- Given that we are using ABC in a misspecified setting
- **ABC inference** is adversely affected by misspecification **[Frazier et al., 2020]**
  - ABC posterior does concentrate all mass on an appropriately defined pseudo-true value $\theta^\star$

$$\theta^\star = \arg\min_{\theta \in \Theta} d\{b_0, b(\theta)\},$$

  - But, the posterior does not concentrate in a Gaussian manner
  - asymptotic shape of the ABC posterior is non-standard and it's credible sets can have arbitrary coverage

Theoretical Validation

- Given that we are using ABC in a misspecified setting
- **ABC inference** is adversely affected by misspecification **[Frazier et al., 2020]**
  - ABC posterior does concentrate all mass on an appropriately defined pseudo-true value $\theta^\star$

$$\theta^\star = \arg\min_{\theta \in \Theta} d\{b_0, b(\theta)\},$$

  - But, the posterior does not concentrate in a Gaussian manner
  - asymptotic shape of the ABC posterior is non-standard and it's credible sets can have arbitrary coverage
- **Coherent** (loss-based) predictions seem to result despite this

Theoretical Validation

- Consider

Theoretical Validation

- Consider
  - $P_\theta^{(t)}$ as the assumed predictive model

Theoretical Validation

- Consider
  - $P_{\theta}^{(t)}$ as the assumed predictive model
  - $P_{0}^{(t)}$ as the true predictive distribution

Theoretical Validation

- Consider
  - $P_{\theta}^{(t)}$ as the assumed predictive model
  - $P_{0}^{(t)}$ as the true predictive distribution
- We are interested in showing

$$\mathbb{S}_j\left(P_{\theta_j}^{(t)}, P_0^{(t)}\right) \geq \mathbb{S}_j\left(P_{\theta_i}^{(t)}, P_0^{(t)}\right), \quad \forall \theta \in \Theta \text{ and } i \neq j$$

where

- $\mathbb{S}_j\left(\cdot, P_0^{(t)}\right)$: expected score under the true predictive $P_0^{(t)}$ based on a particular scoring rule j

Theoretical Validation

- Consider
  - $P_{\theta}^{(t)}$ as the assumed predictive model
  - $P_{0}^{(t)}$ as the true predictive distribution

- We are interested in showing

$$\mathbb{S}_j\left(P_{\theta_j}^{(t)}, P_0^{(t)}\right) \geq \mathbb{S}_j\left(P_{\theta_i}^{(t)}, P_0^{(t)}\right), \quad \forall \theta \in \Theta \text{ and } i \neq j$$

  where
  - $\mathbb{S}_j\left(\cdot, P_0^{(t)}\right)$: expected score under the true predictive $P_0^{(t)}$ based on a particular scoring rule j

- Problem in our context!

Theoretical Validation

- Consider
  - $P_\theta^{(t)}$ as the assumed predictive model
  - $P_0^{(t)}$ as the true predictive distribution

- We are interested in showing

$$\mathbb{S}_j\left(P_{\theta_j}^{(t)}, P_0^{(t)}\right) \geq \mathbb{S}_j\left(P_{\theta_i}^{(t)}, P_0^{(t)}\right), \quad \forall\theta \in \Theta \text{ and } i \neq j$$

where
  - $\mathbb{S}_j\left(\cdot, P_0^{(t)}\right)$: expected score under the true predictive $P_0^{(t)}$ based on a particular scoring rule j

- Problem in our context!
  - due to the third layer of models: **Auxiliary model**

Empirical Illustration

Empirical Illustration

- Applying **loss-based ABF** to daily returns data on the S & P500 index

Empirical Illustration

- Applying **loss-based ABF** to daily returns data on the S & P500 index
- **Assumed predictive class**:

Empirical Illustration

- Applying **loss-based ABF** to daily returns data on the S & P500 index
- **Assumed predictive class**:
  - **SV model with α-Stable errors** for a continuous financial return, y_t,

$$y_t = \exp\left(\frac{h_t}{2}\right)\varepsilon_t \quad ; \quad \varepsilon_t \overset{iid}{\sim} N(0,1)$$

$$h_t = \omega + \rho h_{t-1} + \sigma_h \eta_t \quad ; \quad \eta_t \overset{iid}{\sim} S(\alpha, -1, 0, dt = 1)$$

Empirical Illustration

- Applying **loss-based ABF** to daily returns data on the S & P500 index
- **Assumed predictive class**:
  - **SV model with α-Stable errors** for a continuous financial return, y_t,

$$y_t = \exp\left(\frac{h_t}{2}\right)\varepsilon_t \quad ; \quad \varepsilon_t \overset{iid}{\sim} N(0,1)$$

$$h_t = \omega + \rho h_{t-1} + \sigma_h \eta_t \quad ; \quad \eta_t \overset{iid}{\sim} S(\alpha, -1, 0, dt = 1)$$

  - **Transition density is unavailable**

Empirical Illustration

- Applying **loss-based ABF** to daily returns data on the S & P500 index
- **Assumed predictive class**:
  - **SV model with α-Stable errors** for a continuous financial return, y_t,

$$y_t = \exp\left(\frac{h_t}{2}\right)\varepsilon_t \quad ; \quad \varepsilon_t \overset{iid}{\sim} N(0,1)$$

$$h_t = \omega + \rho h_{t-1} + \sigma_h \eta_t \quad ; \quad \eta_t \overset{iid}{\sim} S(\alpha, -1, 0, dt = 1)$$

  - **Transition density is unavailable**
- $\theta = (\omega, \rho, \sigma_h^2, \alpha)'$

Empirical Illustration: Results

Empirical Illustration: Results

ABC Score			Average out-of-sample score		
	LS	$CS_{<10\%}$	$CS_{<20\%}$	$CS_{>80\%}$	$CS_{>90\%}$
LS	*3.3967*	0.0268	0.2923	0.5170	0.1523
$CS_{<10\%}$	3.4058	**0.0391**	0.3048	0.5122	0.1500
$CS_{<20\%}$	3.4053	0.0383	*0.3041*	0.5126	0.1502
$CS_{>80\%}$	3.3772	0.0052	0.2713	**0.5192**	0.1547
$CS_{>90\%}$	3.3849	0.0134	0.2790	0.5191	*0.1540*

Empirical Illustration: Results

	Average out-of-sample score				
ABC Score	LS	$CS_{<10\%}$	$CS_{<20\%}$	$CS_{>80\%}$	$CS_{>90\%}$
LS	*3.3967*	0.0268	0.2923	0.5170	0.1523
$CS_{<10\%}$	3.4058	**0.0391**	0.3048	0.5122	0.1500
$CS_{<20\%}$	3.4053	0.0383	*0.3041*	0.5126	0.1502
$CS_{>80\%}$	3.3772	0.0052	0.2713	**0.5192**	0.1547
$CS_{>90\%}$	3.3849	0.0134	0.2790	0.5191	*0.1540*

- **Coherent predictions** are in evidence

Thank you!

References I

Bissiri, P. G., Holmes, C. C., and Walker, S. G. (2016).
A general framework for updating belief distributions.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 78(5):1103–1130.

Creel, M. and Kristensen, D. (2015).
ABC of SV: Limited information likelihood inference in stochastic volatility jump-diffusion models.
Journal of Empirical Finance, 31:85–108.

Dean, T. A., Singh, S. S., Jasra, A., and Peters, G. W. (2014).
Parameter estimation for hidden Markov models with intractable likelihoods.
Scandinavian Journal of Statistics, 41(4):970–987.

Drovandi, C. C., Pettitt, A. N., and Faddy, M. J. (2011).
Approximate bayesian computation using indirect inference.
Journal of the Royal Statistical Society: Series C (Applied Statistics), 60(3):317–337.

Drovandi, C. C., Pettitt, A. N., and Lee, A. (2015).
Bayesian indirect inference using a parametric auxiliary model.
Statistical Science, 30(1):72–95.

Frazier, D. T., Loaiza-Maya, R., and Martin, G. M. (2022).
Variational bayes in state space models: Inferential and predictive accuracy.
arXiv preprint arXiv:2106.12262.
Forthcoming, Journal of Computational and Graphical Statistics.

Frazier, D. T., Loaiza-Maya, R., Martin, G. M., and Koo, B. (2021).
Loss-based variational Bayes prediction.
arXiv preprint arXiv:2104.14054.

References II

Frazier, D. T., Maneesoonthorn, W., Martin, G. M., and McCabe, B. P. (2019).
Approximate bayesian forecasting.
International Journal of Forecasting, 35(2):521–539.

Frazier, D. T., Martin, G. M., Robert, C. P., and Rousseau, J. (2018).
Asymptotic properties of approximate bayesian computation.
Biometrika, 105(3):593–607.

Frazier, D. T., Robert, C. P., and Rousseau, J. (2020).
Model misspecification in approximate bayesian computation: consequences and diagnostics.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82(2):421–444.

Giummolè, F., Mameli, V., Ruli, E., and Ventura, L. (2019).
Objective bayesian inference with proper scoring rules.
Test, 28(3):728–755.

Gneiting, T. and Raftery, A. E. (2007).
Strictly proper scoring rules, prediction, and estimation.
Journal of the American statistical Association, 102(477):359–378.

Knoblauch, J., Jewson, J., and Damoulas, T. (2019).
Generalized variational inference: Three arguments for deriving new posteriors.
arXiv preprint arXiv:1904.02063.

Loaiza-Maya, R., Martin, G. M., and Frazier, D. T. (2021a).
Focused bayesian prediction.
Journal of Applied Econometrics, 36(5):517–543.

References III

Loaiza-Maya, R., Smith, M. S., Nott, D. J., and Danaher, P. J. (2021b).
Fast and accurate variational inference for models with many latent variables.
Forthcoming. Journal of Econometrics.

Martin, G. M., McCabe, B. P., Frazier, D. T., Maneesoonthorn, W., and Robert, C. P. (2019).
Auxiliary likelihood-based approximate bayesian computation in state space models.
Journal of Computational and Graphical Statistics, 28(3):508–522.

Pacchiardi, L. and Dutta, R. (2021).
Generalized bayesian likelihood-free inference using scoring rules estimators.
arXiv preprint arXiv:2104.03889.

Quiroz, M., Nott, D. J., and Kohn, R. (2022).
Gaussian variational approximation for high-dimensional state space models.
https://arXiv:1801.07873.
Forthcoming, Bayesian Analysis.

Tran, M.-N., Nott, D. J., and Kohn, R. (2017).
Variational Bayes with intractable likelihood.
Journal of Computational and Graphical Statistics, 26(4):873–882.