Forecasting with Approximate Bayesian Computation (ABC)

Gael Martin

Monash University, Melbourne

IMS Workshop

August, 2018
Under two scenarios.....
Under two scenarios.....

1. Under **correct** specification of the **data generating process (DGP):**
Under correct specification of the data generating process (DGP):

- Frazier, Maneesoonthorn, Martin and McCabe, 2018
Under two scenarios.....

1. Under **correct** specification of the **data generating process** (DGP):

   - Frazier, Maneesoonthorn, Martin and McCabe, 2018
   - *Approximate Bayesian Forecasting* (‘ABF’) *In Press*
Under two scenarios.....

1. Under **correct** specification of the **data generating process** (DGP):
   - Frazier, Maneesoonthorn, Martin and McCabe, 2018
   - *Approximate Bayesian Forecasting (‘ABF’) (In Press)*

2. Under **misspecification** of the DGP:
Under two scenarios.....

1. Under **correct** specification of the **data generating process** (DGP):
   - Frazier, Maneesoonthorn, Martin and McCabe, 2018
   - *Approximate Bayesian Forecasting (‘ABF’) (In Press)*

2. Under **misspecification** of the DGP:
   - Frazier and Martin, 2018
Under two scenarios.....

1. Under **correct** specification of the **data generating process** (DGP):
   - Frazier, Maneesoonthorn, Martin and McCabe, 2018
   - *Approximate Bayesian Forecasting* (‘ABF’) *In Press*

2. Under **mis specification** of the **DGP**:
   - Frazier and Martin, 2018:
     - Very preliminary!!
Exact Bayesian forecasting

Distribution of interest is:

\[ p_{\text{exact}}(y_{T+1}|y_j) = \int p(y_{T+1}|\theta, y_j) p(\theta|y_j) d\theta = E_{\theta|y_j}[p(y_{T+1}|\theta, y_j)] \]

Exact (marginal) predictive = expectation of the conditional predictive

Conditional predictive reflects the assumed DGP (on which \( p(\theta|y_j) \) is also based)
Exact Bayesian forecasting

- Distribution of interest is:

\[ p_{\text{exact}}(y_{T+1} | y) \]
Distribution of interest is:

\[ p_{\text{exact}}(y_{T+1}|y) \]
Distribution of interest is:

\[ p_{\text{exact}}(y_{T+1}|y) = \int_{\theta} p(y_{T+1}, \theta|y) d\theta \]
Exact Bayesian forecasting

- Distribution of interest is:

\[
p_{\text{exact}}(y_{T+1}|y) = \int_{\theta} p(y_{T+1}, \theta|y) d\theta
\]

\[
= \int_{\theta} p(y_{T+1}|y, \theta)p(\theta|y) d\theta
\]
Distribution of interest is:

\[
p_{\text{exact}}(y_{T+1}|y) = \int_{\theta} p(y_{T+1}, \theta|y) d\theta \\
= \int_{\theta} p(y_{T+1}|y, \theta)p(\theta|y) d\theta \\
= E_{\theta|y}[p(y_{T+1}|y, \theta)]
\]
Distribution of interest is:

\[ p_{\text{exact}}(y_{T+1}|y) = \int_{\theta} p(y_{T+1}, \theta|y) d\theta \]

\[ = \int_{\theta} p(y_{T+1}|y, \theta)p(\theta|y) d\theta \]

\[ = E_{\theta|y}[p(y_{T+1}|y, \theta)] \]

**Exact (marginal) predictive** = expectation of the conditional predictive
Exact Bayesian forecasting

- Distribution of interest is:

\[ p_{\text{exact}}(y_{T+1}|y) = \int_\theta p(y_{T+1}, \theta|y) d\theta \]

\[ = \int_\theta p(y_{T+1}|y, \theta) p(\theta|y) d\theta \]

\[ = E_{\theta|y} [p(y_{T+1}|y, \theta)] \]

- **Exact (marginal) predictive** = expectation of the **conditional** predictive

- **Conditional** predictive reflects the **assumed** DGP
Exact Bayesian forecasting

Distribution of interest is:

\[ p_{\text{exact}}(y_{T+1}|y) = \int_{\theta} p(y_{T+1}, \theta|y) d\theta \]

\[ = \int_{\theta} p(y_{T+1}|y, \theta)p(\theta|y) d\theta \]

\[ = E_{\theta|y} [p(y_{T+1}|y, \theta)] \]

**Exact (marginal) predictive** = expectation of the **conditional** predictive

**Conditional** predictive reflects the **assumed DGP**

(on which \( p(\theta|y) \) is also based)
Given $M$ draws from $p(\theta_j | y)$ (via a Markov chain Monte Carlo algorithm, say) $p_{\text{exact}}(y_{T+1} | j, y)$ can be estimated as either:

$$p_{\text{exact}}(y_{T+1} | j, y) = \frac{1}{M} \sum_{i=1}^{M} p(y_{T+1} | j, y_i, \theta(i))$$

or:

$$p_{\text{exact}}(y_{T+1} | j, y) \text{ constructed from draws of } y_i \text{ simulated from } p(y_{T+1} | j, y, \theta(i))$$

e.g. MCMC exact Bayesian forecasting (up to simulation error)
Exact Bayesian forecasting

- Given $M$ draws from $p(\theta|y)$ (via a Markov chain Monte Carlo algorithm, say)
Given $M$ draws from $p(\theta|y)$ (via a Markov chain Monte Carlo algorithm, say)
Exact Bayesian forecasting

- Given $M$ draws from $p(\theta|y)$ (via a Markov chain Monte Carlo algorithm, say)

- $p_{\text{exact}}(y_{T+1}|y)$ can be estimated as
Exact Bayesian forecasting

- Given \( M \) draws from \( p(\theta|y) \) (via a Markov chain Monte Carlo algorithm, say)

- \( p_{\text{exact}}(y_{T+1}|y) \) can be estimated as

  1. either:

     \[
     \hat{p}_{\text{exact}}(y_{T+1}|y) = \frac{1}{M} \sum_{i=1}^{M} p(y_{T+1}|y, \theta^{(i)})
     \]
Exact Bayesian forecasting

- Given $M$ draws from $p(\theta | y)$ (via a Markov chain Monte Carlo algorithm, say)
- $p_{\text{exact}}(y_{T+1} | y)$ can be estimated as
  1. either:
     \[ \hat{p}_{\text{exact}}(y_{T+1} | y) = \frac{1}{M} \sum_{i=1}^{M} p(y_{T+1} | y, \theta^{(i)}) \]
  2. or: $\hat{p}_{\text{exact}}(y_{T+1} | y)$ constructed from draws of $y_{T+1}^{(i)}$ simulated from $p(y_{T+1} | y, \theta^{(i)})$
Given $M$ draws from $p(\theta|y)$ (via a Markov chain Monte Carlo algorithm, say)

$p_{\text{exact}}(y_{T+1}|y)$ can be estimated as

1. either:

$$\widehat{p}_{\text{exact}}(y_{T+1}|y) = \frac{1}{M} \sum_{i=1}^{M} p(y_{T+1}|y, \theta^{(i)})$$

2. or: $p_{\text{exact}}(y_{T+1}|y)$ constructed from draws of $y_{T+1}^{(i)}$ simulated from $p(y_{T+1}|y, \theta^{(i)})$

i.e. MCMC $\Rightarrow$ exact Bayesian forecasting
Exact Bayesian forecasting

- Given $M$ draws from $p(\theta|y)$ (via a Markov chain Monte Carlo algorithm, say)

- $p_{exact}(y_{T+1}|y)$ can be estimated as
  1. either:
     
     $$
     \hat{p}_{exact}(y_{T+1}|y) = \frac{1}{M} \sum_{i=1}^{M} p(y_{T+1}|y, \theta^{(i)})
     $$
  2. or: $p_{exact}(y_{T+1}|y)$ constructed from draws of $y^{(i)}_{T+1}$ simulated from $p(y_{T+1}|y, \theta^{(i)})$

- i.e. MCMC $\Rightarrow$ exact Bayesian forecasting

  (up to simulation error)
How to conduct Bayesian forecasting when $p(\theta|y)$ is inaccessible?

- Either because the assumed DGP $p(y|\theta)$ is intractable in the sense that (parts of) the DGP unavailable in closed form.
- Or when the dimension of $\theta$ so large that exploration of $p(\theta|y)$ via exact methods is deemed to be too computationally burdensome.
- Or there is insufficient expertise to structure an efficient MCMC algorithm.
How to conduct Bayesian forecasting when \( p(\theta|y) \) is inaccessible?

\[ \Rightarrow \text{draws from it are unavailable} \]
Approximate Bayesian forecasting

- How to conduct Bayesian forecasting when $p(\theta|y)$ is inaccessible?
- $\Rightarrow$ draws from it are unavailable
- **Either** because the assumed DGP $p(y|\theta)$ is **intractable**
How to conduct Bayesian forecasting when $p(\theta|y)$ is inaccessible?

$\Rightarrow$ draws from it are unavailable

Either because the assumed DGP $p(y|\theta)$ is intractable

in the sense that (parts of) the DGP unavailable in closed form
How to conduct Bayesian forecasting when \( p(\theta|y) \) is inaccessible?

\[ \Rightarrow \text{draws from it are unavailable} \]

Either because the assumed DGP \( p(y|\theta) \) is intractable

in the sense that (parts of) the DGP unavailable in closed form

Or when the dimension of \( \theta \) so large
Approximate Bayesian forecasting

- How to conduct Bayesian forecasting when $p(\theta | y)$ is inaccessible?
- $\Rightarrow$ draws from it are unavailable
- **Either** because the assumed DGP $p(y | \theta)$ is intractable
  - in the sense that (parts of) the DGP unavailable in closed form
- **Or** when the dimension of $\theta$ so large
  - that exploration of $p(\theta | y)$ via **exact** methods is deemed to be too computationally burdensome
Approximate Bayesian forecasting

- How to conduct Bayesian forecasting when $p(\theta|y)$ is inaccessible?
  - \[ \Rightarrow \text{draws from it are unavailable} \]
  - **Either** because the assumed DGP $p(y|\theta)$ is **intractable**
  - in the sense that (parts of) the DGP **unavailable** in closed form
  - **Or** when the dimension of $\theta$ so large
  - that exploration of $p(\theta|y)$ via **exact** methods is deemed to be too computationally burdensome
  - **Or** there is insufficient expertise to structure an efficient MCMC algorithm
Can/must resort to **approximate Bayesian inference**
Approximate Bayesian forecasting

- Can/must resort to approximate Bayesian inference

  \[ \Rightarrow \text{goal then is to produce an approximation to } p(\theta | y) \]
Approximate Bayesian forecasting

- Can/must resort to approximate Bayesian inference
- ⇒ goal then is to produce an approximation to $p(\theta|y)$
- ⇒ an approximation to $p_{\text{exact}}(y_{T+1}|y)$
Can/must resort to **approximate Bayesian inference**

$$\Rightarrow \textbf{goal} \text{ then is to produce an approximation to } p(\theta|y)$$

$$\Rightarrow \text{an approximation to } p_{\text{exact}}(y_{T+1}|y)$$

**Approximations to** $p(\theta|y)$?
Can/must resort to approximate Bayesian inference

⇒ goal then is to produce an approximation to $p(\theta|y)$

⇒ an approximation to $p_{\text{exact}}(y_{T+1}|y)$

Approximations to $p(\theta|y)$?

- Variational Bayes
Can/must resort to approximate Bayesian inference

⇒ goal then is to produce an approximation to $p(\theta|y)$

⇒ an approximation to $p_{\text{exact}}(y_{T+1}|y)$

Approximations to $p(\theta|y)$?

- Variational Bayes
- Integrated nested Laplace (INLA)
Approximate Bayesian forecasting

- Can/must resort to **approximate Bayesian inference**
- \( \Rightarrow \) **goal** then is to produce an **approximation to** \( p(\theta|y) \)
- \( \Rightarrow \) an **approximation to** \( p_{\text{exact}}(y_{T+1}|y) \)

**Approximations to** \( p(\theta|y) \)?

- Variational Bayes
- Integrated nested Laplace (INLA)
- Synthetic likelihood
Approximate Bayesian forecasting

- Can/must resort to approximate Bayesian inference

  ⇒ goal then is to produce an approximation to $p(\theta|y)$

  ⇒ an approximation to $p_{exact}(y_{T+1}|y)$

- Approximations to $p(\theta|y)$?
  - Variational Bayes
  - Integrated nested Laplace (INLA)
  - Synthetic likelihood
  - Approximate Bayesian computation (ABC)
Approximate Bayesian forecasting

- Can/must resort to approximate Bayesian inference

  ⇒ goal then is to produce an approximation to \( p(\theta|y) \)

  ⇒ an approximation to \( p_{\text{exact}}(y_{T+1}|y) \)

- Approximations to \( p(\theta|y) \)?
  - Variational Bayes
  - Integrated nested Laplace (INLA)
  - Synthetic likelihood
  - Approximate Bayesian computation (ABC)

- All of which could be viewed as yielding ‘approximate Bayesian forecasting’
Approximate Bayesian forecasting

- Can/must resort to **approximate Bayesian inference**
- \( \Rightarrow \) **goal** then is to produce **an approximation to** \( p(\theta|y) \)
- \( \Rightarrow \) an **approximation to** \( p_{\text{exact}}(y_{T+1}|y) \)

**Approximations to** \( p(\theta|y) \)?

- Variational Bayes
- Integrated nested Laplace (INLA)
- Synthetic likelihood
- Approximate Bayesian computation (ABC)

- All of which could be viewed as yielding ‘**approximate Bayesian forecasting**’

- **Our focus is on ABC**
ABC (basic form) in a nut shell!

Aim is to produce draws from an approximation to $p(\theta_jy)$ and use draws to estimate that approximation.

The simplest (accept/reject) form of the algorithm:

1. Simulate $i_1, i_2, \ldots, i_N$, draws of $\theta_i$ from $p(\theta)$.

2. Simulate pseudo-data $z_{i_1}, i_2, \ldots, z_{i_N}$, from $p(z_j|\theta_i)$.

3. Select $\theta_i$ such that:

$$df(\eta(y), \eta(z_{i_i})) \epsilon \eta(\cdot)$$

is a (vector) summary statistic $df$. $g$ is a distance criterion, the tolerance $\epsilon$ is arbitrarily small.

4. Selected draws (simulation-based estimate of $p(\theta_j|\eta(y))$).
ABC (basic form) in a nut shell!

- Aim is to produce **draws** from an **approximation** to $p(\theta|y)$
ABC (basic form) in a nut shell!

- Aim is to produce **draws** from an **approximation** to $p(\theta|y)$
- and use draws to **estimate** that **approximation**
ABC (basic form) in a nutshell!

- Aim is to produce **draws** from an **approximation** to \( p(\theta|y) \).
- and use draws to **estimate** that **approximation**
- The simplest (accept/reject) form of the algorithm:

\[
1. \text{Simulate } i = 1, 2, \ldots, N, i. \text{ draws of } \theta_i \text{ from } p(\theta)
\]

\[
2. \text{Simulate pseudo-data } z_i, i = 1, 2, \ldots, N, \text{ from } p(z|\theta_i)
\]

\[
3. \text{Select } \theta_i \text{ such that: } d_f \eta(y), \eta(z_i) \text{ is a (vector) summary statistic } d_f, g \text{ is a distance criterion}\]

\[
\text{the tolerance } \varepsilon \text{ is arbitrarily small}
\]

\[
4. \text{Selected draws (simulation-based estimate of } p(\theta_j \eta(y))
\]
ABC (basic form) in a nut shell!

- Aim is to produce **draws** from an **approximation** to $p(\theta | y)$
- and use draws to **estimate** that **approximation**
- The simplest (accept/reject) form of the algorithm:
  1. Simulate $i = 1, 2, ..., N$, i.i.d. draws of $\theta^i$ from $p(\theta)$
ABC (basic form) in a nut shell!

- Aim is to produce **draws** from an **approximation** to \( p(\theta | y) \)
- and use draws to **estimate** that approximation
- The simplest (accept/reject) form of the algorithm:
  1. Simulate \( i = 1, 2, ..., N, \) \( i.i.d. \) draws of \( \theta^i \) from \( p(\theta) \)
  2. Simulate **pseudo-data** \( z^i, i = 1, 2, ..., N, \) from \( p(z | \theta^i) \)
ABC (basic form) in a nut shell!

- Aim is to produce draws from an approximation to $p(\theta | y)$
- and use draws to estimate that approximation
- The simplest (accept/reject) form of the algorithm:
  1. Simulate $i = 1, 2, ..., N$, i.i.d. draws of $\theta^i$ from $p(\theta)$
  2. Simulate pseudo-data $z^i$, $i = 1, 2, ..., N$, from $p(z | \theta^i)$
  3. Select $\theta^i$ such that:

\[
d\{\eta(y), \eta(z^i)\} \leq \varepsilon
\]
ABC (basic form) in a nut shell!

- Aim is to produce **draws** from an **approximation** to $p(\theta|y)$
- and use draws to **estimate** that approximation

The simplest (accept/reject) form of the algorithm:

1. Simulate $i = 1, 2, \ldots, N$, i.i.d. draws of $\theta^i$ from $p(\theta)$
2. Simulate **pseudo-data** $z^i$, $i = 1, 2, \ldots, N$, from $p(z|\theta^i)$
3. Select $\theta^i$ such that:

$$d\{\eta(y), \eta(z^i)\} \leq \varepsilon$$

- $\eta(.)$ is a (vector) **summary statistic**
ABC (basic form) in a nut shell!

- Aim is to produce **draws** from an **approximation** to $p(\theta|y)$
- and use draws to **estimate** that **approximation**
- The simplest (accept/reject) form of the algorithm:

1. Simulate $i = 1, 2, ..., N$, i.i.d. draws of $\theta^i$ from $p(\theta)$
2. Simulate **pseudo-data** $z^i$, $i = 1, 2, ..., N$, from $p(z|\theta^i)$
3. Select $\theta^i$ such that:

$$d\{\eta(y), \eta(z^i)\} \leq \varepsilon$$

- $\eta(.)$ is a (vector) **summary statistic**
- $d\{.\}$ is a distance criterion
ABC (basic form) in a nut shell!

- Aim is to produce **draws** from an **approximation** to $p(\theta|y)$
- and use draws to **estimate** that **approximation**

The simplest (accept/reject) form of the algorithm:

1. Simulate $i = 1, 2, ..., N$, *i.i.d.* draws of $\theta^i$ from $p(\theta)$
2. Simulate **pseudo-data** $z^i$, $i = 1, 2, ..., N$, from $p(z|\theta^i)$
3. Select $\theta^i$ such that:
   \[
   d\{\eta(y), \eta(z^i)\} \leq \varepsilon
   \]
   - $\eta(.)$ is a (vector) **summary statistic**
   - $d\{.\}$ is a distance criterion
   - the tolerance $\varepsilon$ is arbitrarily small
ABC (basic form) in a nut shell!

- Aim is to produce \textbf{draws} from an \textbf{approximation} to $p(\theta|y)$
- and use draws to \textbf{estimate} that \textbf{approximation}
- The simplest (accept/reject) form of the algorithm:
  1. Simulate $i = 1, 2, ..., N$, \textit{i.i.d.} draws of $\theta^i$ from $p(\theta)$
  2. Simulate \textbf{pseudo-data} $z^i$, $i = 1, 2, ..., N$, from $p(z|\theta^i)$
  3. Select $\theta^i$ such that:

$$d\{\eta(y), \eta(z^i)\} \leq \epsilon$$

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

4. Selected draws $\Rightarrow$ simulation-based estimate of $p(\theta|\eta(y))$
Approximate Bayesian forecasting

- Use draws from $p(\theta | \eta(y))$ to estimate:

$$p_{ABC}(y_{T+1} | y) = \int p(y_{T+1} | y, \theta) p(\theta | \eta(y)) d\theta$$
Approximate Bayesian forecasting

- Use draws from \( p(\theta | \eta(y)) \) to estimate:

\[
p_{ABC}(y_{T+1} | y) = \int p(y_{T+1} | y, \theta) p(\theta | \eta(y)) d\theta
\]
Use draws from $p(\theta | \eta(y))$ to estimate:

$$p_{ABC}(y_{T+1} | y) = \int p(y_{T+1} | y, \theta) p(\theta | \eta(y)) d\theta$$

$$= \text{an ‘approximate Bayesian predictive’}$$
Approximate Bayesian forecasting

- Use draws from $p(\theta|\eta(y))$ to estimate:

$$p_{ABC}(y_{T+1}|y) = \int p(y_{T+1}|y, \theta)p(\theta|\eta(y))d\theta$$

= an 'approximate Bayesian predictive’

- What is $p_{ABC}(y_{T+1}|y)$ & how does it relate to $p_{exact}(y_{T+1}|y)$?
Approximate Bayesian forecasting

- Use draws from $p(\theta|\eta(y))$ to estimate:

$$p_{ABC}(y_{T+1}|y) = \int p(y_{T+1}|y, \theta)p(\theta|\eta(y))d\theta = \text{an ‘approximate Bayesian predictive’}$$

- What is $p_{ABC}(y_{T+1}|y)$ & how does it relate to $p_{exact}(y_{T+1}|y)$?

- We show (in ‘ABF’, 2018) that:
Approximate Bayesian forecasting

- Use draws from $p(\theta | \eta(y))$ to estimate:

$$p_{ABC}(y_{T+1}|y) = \int p(y_{T+1}|y, \theta) p(\theta | \eta(y)) d\theta$$

= an ‘approximate Bayesian predictive’

- What is $p_{ABC}(y_{T+1}|y)$ & how does it relate to $p_{exact}(y_{T+1}|y)$?

- We show (in ‘ABF’, 2018) that:

  - $p_{ABC}(y_{T+1}|y)$ is a proper density function
Approximate Bayesian forecasting

- Use draws from $p(\theta|\eta(y))$ to estimate:

$$
p_{ABC}(y_{T+1}|y) = \int p(y_{T+1}|y, \theta)p(\theta|\eta(y))d\theta
= \text{an ‘approximate Bayesian predictive’}
$$

- What is $p_{ABC}(y_{T+1}|y)$ & how does it relate to $p_{exact}(y_{T+1}|y)$?

- We show (in ‘ABF’, 2018) that:
  - $p_{ABC}(y_{T+1}|y)$ is a proper density function
  - $p_{ABC}(y_{T+1}|y) = p_{exact}(y_{T+1}|y)$ iff $\eta(y)$ is sufficient (!)
Approximate Bayesian forecasting

- Use draws from $p(\theta|\eta(y))$ to estimate:

$$p_{ABC}(y_{T+1}|y) = \int p(y_{T+1}|y, \theta)p(\theta|\eta(y))d\theta$$

= an ‘approximate Bayesian predictive’

- What is $p_{ABC}(y_{T+1}|y)$ & how does it relate to $p_{exact}(y_{T+1}|y)$?

- We show (in ‘ABF’, 2018) that:
  - $p_{ABC}(y_{T+1}|y)$ is a **proper** density function
  - $p_{ABC}(y_{T+1}|y) = p_{exact}(y_{T+1}|y)$ iff $\eta(y)$ is **sufficient** (!)
  - $p_{ABC}(y_{T+1}|y) \approx p_{exact}(y_{T+1}|y)$ even when $\eta(y)$ is **not** sufficient
Furthermore......

Under Bayesian consistency of:

\[ p(\theta_j y) \] (standard regularity) and

\[ p(\theta_j \eta(y)) \] (Frazier, Martin, Robert and Rousseau, 2018)

the predictive distributions:

\[ P_{\text{exact}} \] and \[ P_{\text{ABC}} \] 'merge', in the sense that:

\[ \rho_{TV}(P_{\text{exact}}, P_{\text{ABC}}) = \sup_B \left( P_{\text{exact}}(B) - P_{\text{ABC}}(B) \right) \] (Blackwell and Dubins, 1962)

for large enough \( T_{\text{exact}} \) and (consistent) ABC-based predictives are equivalent!

Gael Martin, Monash University, Melbourne. Forecasting with Approximate Bayesian Comp
Furthermore......

- Under **Bayesian consistency** of:

\[ p(\theta) \]

(standard regularity) and

\[ p(\theta \eta | y) \] (Frazier, Martin, Robert and Rousseau, 2018)

the predictive distributions:

\[ P_{\text{exact}} \] and \[ P_{\text{ABC}} \] ‘merge’ in the sense that:

\[ \rho_{TV}\left(P_{\text{exact}}, P_{\text{ABC}}\right) = \sup_{B \in F} \left\{ \int_{B} P_{\text{exact}}(B) \right\} \int_{B} P_{\text{ABC}}(B) \right\} = o(1) \]

Blackwell and Dubins (1962)

for large enough \( T_{\text{exact}} \) and (consistent) ABC-based predictives are equivalent!
Furthermore......

- Under **Bayesian consistency** of:
  - $p(\theta|y)$ (standard regularity) and
Furthermore......

- **Under Bayesian consistency** of:
  - $p(\theta|y)$ (standard regularity) and
  - $p(\theta|\eta(y))$ (Frazier, Martin, Robert and Rousseau, 2018)

$P_{\text{exact}}$ and $P_{\text{ABC}}$ merge, in the sense that:

$$\rho_{\text{TV}} f_{P_{\text{exact}}}, P_{\text{ABC}} g = \sup_B 2F_{\text{exact}}(B) P_{\text{exact}}(B) = o_1$$

for large enough $T$ and (consistent) ABC-based predictives are equivalent!
Furthermore......

- Under **Bayesian consistency** of:
  - $p(\theta|y)$ (standard regularity) and
  - $p(\theta|\eta(y))$ (Frazier, Martin, Robert and Rousseau, 2018)

- the predictive distributions:

$$P_{\text{exact}}(\cdot) \text{ and } P_{\text{ABC}}(\cdot)$$
Furthermore......

- Under **Bayesian consistency** of:
  - $p(\theta|y)$ (standard regularity) and
  - $p(\theta|\eta(y))$ (Frazier, Martin, Robert and Rousseau, 2018)

- the predictive distributions:
  
  $$P_{\text{exact}}(\cdot) \text{ and } P_{\text{ABC}}(\cdot)$$
Furthermore......

- **Under Bayesian consistency** of:
  - $p(\theta|y)$ (standard regularity) and
  - $p(\theta|\eta(y))$ (Frazier, Martin, Robert and Rousseau, 2018)

- the predictive distributions:
  - $P_{\text{exact}}(\cdot)$ and $P_{\text{ABC}}(\cdot)$

  ‘merge’, in the sense that:

  $$
  \rho_{TV}\{P_{\text{exact}}, P_{\text{ABC}}\} = \sup_{B \in \mathcal{F}} |P_{\text{exact}}(B) - P_{\text{ABC}}(B)| = o_{\mathbb{P}}(1)
  $$

- Blackwell and Dubins (1962)
Furthermore......

- **Under Bayesian consistency of:**
  - $p(\theta|y)$ (standard regularity) and
  - $p(\theta|\eta(y))$ (*Frazier, Martin, Robert and Rousseau, 2018*)

- the predictive distributions:
  \[ P_{exact}(\cdot) \text{ and } P_{ABC}(\cdot) \]

  ‘merge’, in the sense that:
  \[
  \rho_{TV}\left\{P_{exact}, P_{ABC}\right\} = \sup_{B \in \mathcal{F}} \left| P_{exact}(B) - P_{ABC}(B) \right| = o_{P}(1)
  \]

- **Blackwell and Dubins (1962)**

- $\Rightarrow$ for large enough $T$ exact and (consistent) ABC-based predictives are equivalent!
Furthermore...

- Under **asymptotic normality** of:

\[ p(\theta_j y) \] (standard regularity) and \[ p(\theta_j \eta(y)) \] (Frazier, Martin, Robert and Rousseau, 2018) inequality result regarding the predictive accuracy of \[ p_{\text{exact}}(y_T + 1|y) \] vs \[ p_{\text{ABC}}(y_T + 1|y) \] using a proper scoring rule:

\[ E[S(p_{\text{exact}}, y_{T+1})] = \int_{\Omega} S(p_{\text{exact}}, y_{T+1}) p(y_{T+1}|\theta_0) \, dy \]

\[ = E[S(p_{\text{ABC}}, y_{T+1})] \]

Gael Martin, Monash University, Melbourne. Forecasting with Approximate Bayesian Comp
Furthermore......

- Under **asymptotic normality** of:
  - $p(\theta|y)$ (standard regularity) and
Furthermore......

- Under **asymptotic normality** of:
  
  1. $p(\theta|y)$ (standard regularity) and
  2. $p(\theta|\eta(y))$ (Frazier, Martin, Robert and Rousseau, 2018)
Furthermore......

- Under **asymptotic normality** of:
  - $p(\theta|y)$ (standard regularity) and
  - $p(\theta|\eta(y))$ (Frazier, Martin, Robert and Rousseau, 2018)
- $\Rightarrow$ **inequality** result regarding the **predictive accuracy** of
  $p_{\text{exact}}(y_{T+1}|y)$ vs $p_{\text{ABC}}(y_{T+1}|y)$
Furthermore......

- Under **asymptotic normality** of:
  - $p(\theta|y)$ (standard regularity) and
  - $p(\theta|\eta(y))$ (Frazier, Martin, Robert and Rousseau, 2018)

- ⇒ **inequality** result regarding the **predictive accuracy** of
  - $p_{\text{exact}}(y_{T+1}|y)$ vs $p_{ABC}(y_{T+1}|y)$

- using a **proper scoring rule**: $S(p_{\text{exact}}, y_{T+1})$
Furthermore……

- Under **asymptotic normality** of:
  - $p(\theta|y)$ (standard regularity) and
  - $p(\theta|\eta(y))$ (Frazier, Martin, Robert and Rousseau, 2018)

  $\Rightarrow$ inequality result regarding the **predictive accuracy** of
  $p_{\text{exact}}(y_{T+1}|y)$ vs $p_{\text{ABC}}(y_{T+1}|y)$

- using a **proper scoring rule**: $S(p_{\text{exact}}, y_{T+1})$

  $\Rightarrow$ for large (but finite) $T$:
Furthermore......

- Under **asymptotic normality** of:
  - \( p(\theta|y) \) (standard regularity) and
  - \( p(\theta|\eta(y)) \) (**Frazier, Martin, Robert and Rousseau, 2018**)
  - \( \Rightarrow \) **inequality** result regarding the **predictive accuracy** of
  \( p_{\text{exact}}(y_{T+1}|y) \) vs \( p_{ABC}(y_{T+1}|y) \)
  - using a **proper scoring rule**: \( S(p_{\text{exact}}, y_{T+1}) \)
  - \( \Rightarrow \) for large (but finite) \( T \):

\[ E[S(p_{\text{exact}}, y_{T+1})] = \int_{y_{T+1}} S(p_{\text{exact}}, y_{T+1}) \, \frac{1}{\Omega} \, dy_{T+1} = \int_{y_{T+1}} S(p_{ABC}, y_{T+1}) \, \frac{1}{\Omega} \, dy_{T+1} \]
Furthermore…

- Under **asymptotic normality** of:
  - \( p(\theta|y) \) (standard regularity) and
  - \( p(\theta|\eta(y)) \) (**Frazier, Martin, Robert and Rousseau, 2018**)
  - \( \Rightarrow \) **inequality** result regarding the **predictive accuracy** of
    \( p_{exact}(y_{T+1}|y) \) vs \( p_{ABC}(y_{T+1}|y) \)
  - using a **proper scoring rule**: \( S(p_{exact}, y_{T+1}) \)
  - \( \Rightarrow \) for large (but finite) \( T \):

\[
E[S(p_{exact}, y_{T+1})] = \int_{y \in \Omega} S(p_{exact}, y_{T+1}) \frac{p(y_{T+1}|y, \theta_0)}{p_{truth}} dy_{T+1}
\]
Furthermore......

- **Under asymptotic normality** of:
  - $p(\theta|y)$ (standard regularity) and
  - $p(\theta|\eta(y))$ (*Frazier, Martin, Robert and Rousseau, 2018*)

  \[ \Rightarrow \text{inequality result regarding the predictive accuracy of} \]
  \[ p_{\text{exact}}(y_{T+1}|y) \text{ vs } p_{\text{ABC}}(y_{T+1}|y) \]

- **using a proper scoring rule**: $S(p_{\text{exact}}, y_{T+1})$

  \[ \Rightarrow \text{for large (but finite) } T : \]

  \[
  E[S(p_{\text{exact}}, y_{T+1})] = \int_{y \in \Omega} S(p_{\text{exact}}, y_{T+1}) p(y_{T+1}|y, \theta_0) dy_{T+1} \\
  \geq \int_{y \in \Omega} S(p_{\text{ABC}}, y_{T+1}) p(y_{T+1}|y, \theta_0) dy_{T+1} \\
  = E[S(p_{\text{ABC}}, y_{T+1})]
  \]
Example: MA(2): $T = 500$

Consider (simple) example:

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}$$
Example: MA(2): $T = 500$

- Consider (simple) example:
  \[ y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} \]
- $e_t \sim i.i.d. N(0, \sigma_0)$ with true: $\theta_{10} = 0.8; \theta_{20} = 0.6; \sigma_0 = 1.0$
Example: MA(2): $T = 500$

- Consider (simple) example:
  \[ y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} \]

- \( e_t \sim i.i.d. N(0, \sigma_0) \) with true: \( \theta_{10} = 0.8; \theta_{20} = 0.6; \sigma_0 = 1.0 \)

- Use sample autocovariances
  \[ \gamma_l = \text{cov}(y_t, y_{t-l}) \]
Example: MA(2): $T = 500$

- Consider (simple) example:

  \[ y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} \]

- $e_t \sim i.i.d. N(0, \sigma_0)$ with **true**: $\theta_{10} = 0.8; \theta_{20} = 0.6; \sigma_0 = 1.0$

- Use **sample autocovariances**

  \[ \gamma_l = \text{cov}(y_t, y_{t-l}) \]

- to construct (alternative vectors of) summary statistics:

  \[ \eta^{(1)}(y) = (\gamma_0, \gamma_1)' \]
Example: MA(2): $T = 500$

- Consider (simple) example:
  \[ y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} \]

- $e_t \sim i.i.d.N(0, \sigma_0)$ with true: $\theta_{10} = 0.8; \theta_{20} = 0.6; \sigma_0 = 1.0$

- Use sample autocovariances
  \[ \gamma_l = \text{cov}(y_t, y_{t-l}) \]

- to construct (alternative vectors of) summary statistics:
  \[ \eta^{(1)}(y) = (\gamma_0, \gamma_1)' \]
Example: MA(2): $T = 500$

- Consider (simple) example:
  \[ y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} \]
- \[ e_t \sim i.i.d. N(0, \sigma_0) \text{ with true: } \theta_{10} = 0.8; \theta_{20} = 0.6; \sigma_0 = 1.0 \]
- Use sample autocovariances
  \[ \gamma_l = \text{cov}(y_t, y_{t-l}) \]
- to construct (alternative vectors of) summary statistics:
  \[ \eta^{(1)}(y) = (\gamma_0, \gamma_1)'; \eta^{(2)}(y) = (\gamma_0, \gamma_1, \gamma_2)' \]
Consider (simple) example:

\[ y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} \]

\[ e_t \sim i.i.d. N(0, \sigma_0) \text{ with true: } \theta_{10} = 0.8; \theta_{20} = 0.6; \sigma_0 = 1.0 \]

Use sample autocovariances

\[ \gamma_l = \text{cov}(y_t, y_{t-l}) \]

to construct (alternative vectors of) summary statistics:

\[ \eta^{(1)}(y) = (\gamma_0, \gamma_1)' \]
\[ \eta^{(2)}(y) = (\gamma_0, \gamma_1, \gamma_2)' \]
\[ \eta^{(3)}(y) = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)' \]
Consider (simple) example:

\[ y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} \]

\( e_t \sim i.i.d. \mathcal{N}(0, \sigma_0) \) with true: \( \theta_{10} = 0.8; \theta_{20} = 0.6; \sigma_0 = 1.0 \)

Use sample autocovariances

\[ \gamma_l = \text{cov}(y_t, y_{t-l}) \]

to construct (alternative vectors of) summary statistics:

\[ \eta^{(1)}(y) = (\gamma_0, \gamma_1)'; \eta^{(2)}(y) = (\gamma_0, \gamma_1, \gamma_2)'; \eta^{(3)}(y) = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)'; \eta^{(4)}(y) = (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4)' \]
Example: MA(2): $T = 500$

- Consider (simple) example:
  \[ y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} \]
- $e_t \sim i.i.d. N(0, \sigma_0)$ with true: $\theta_{10} = 0.8; \theta_{20} = 0.6; \sigma_0 = 1.0$
- Use sample autocovariances
  \[ \gamma_l = \text{cov}(y_t, y_{t-l}) \]
- to construct (alternative vectors of) summary statistics:
  \[ \eta^{(1)}(y) = (\gamma_0, \gamma_1)'; \eta^{(2)}(y) = (\gamma_0, \gamma_1, \gamma_2)' \]
  \[ \eta^{(3)}(y) = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)' ; \eta^{(4)}(y) = (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4)' \]
- MA dependence $\Rightarrow$ no reduction to sufficiency possible
Example: MA(2): $T = 500$

- Consider (simple) example:
  
  $$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}$$

- $e_t \sim i.i.d. N(0, \sigma_0)$ with true: $\theta_{10} = 0.8$; $\theta_{20} = 0.6$; $\sigma_0 = 1.0$

- Use sample autocovariances
  
  $$\gamma_l = \text{cov}(y_t, y_{t-l})$$

  to construct (alternative vectors of) summary statistics:
  
  $$\eta^{(1)}(y) = (\gamma_0, \gamma_1)'; \quad \eta^{(2)}(y) = (\gamma_0, \gamma_1, \gamma_2)'$$
  
  $$\eta^{(3)}(y) = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)'; \quad \eta^{(4)}(y) = (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4)'$$

- MA dependence $\Rightarrow$ no reduction to sufficiency possible
  
  $$\Rightarrow p(\theta|\eta^{(j)}(y)) \neq p(\theta|y) \text{ for all } j = 1, 2, 3, 4$$
Example: MA(2): \( T = 500 \)

- Consider (simple) example:
  \[
y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}
  \]
- \( e_t \sim i.i.d. N(0, \sigma_0) \) with true: \( \theta_{10} = 0.8; \theta_{20} = 0.6; \sigma_0 = 1.0 \)
- Use sample autocovariances
  \[
  \gamma_l = \text{cov}(y_t, y_{t-l})
  \]
- to construct (alternative vectors of) summary statistics:
  \[
  \eta^{(1)}(y) = (\gamma_0, \gamma_1)'; \eta^{(2)}(y) = (\gamma_0, \gamma_1, \gamma_2)'
  \]
  \[
  \eta^{(3)}(y) = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)'; \eta^{(4)}(y) = (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4)'
  \]
- MA dependence \( \Rightarrow \) no reduction to sufficiency possible
  \[
  \Rightarrow p(\theta | \eta^{(j)}(y)) \neq p(\theta | y) \text{ for all } j = 1, 2, 3, 4
  \]
- What about \( p_{ABC}(y_{T+1}|y) \text{ versus } p_{\text{exact}}(y_{T+1}|y) \)??
Posterior densities: exact and ABC: $T = 500$
Posterior densities: exact and ABC: $T = 500$

Panel (A)

Panel (B)
Posterior densities: exact and ABC: $T = 500$

- **Remember**: using only the most **basic** version of ABC
Predictive densities: exact and ABC: $T = 500$

For large $T$: the exact and approximate predictives are very similar - for all $\eta_j(y)$.
For large $T$: the exact and approximate predictives are very similar - for all $\eta^{(j)}(y)$!!
Expected scores: exact and ABC: $T = 500$

- **Average predictive scores** over 500 out-of-sample values:
**Expected scores: exact and ABC: \( T = 500 \)**

- **Average predictive scores** over 500 out-of-sample values:

<table>
<thead>
<tr>
<th></th>
<th>( \eta(1) )</th>
<th>( \eta(2) )</th>
<th>( \eta(3) )</th>
<th>( \eta(4) )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LS</strong></td>
<td>-1.43</td>
<td>-1.42</td>
<td>-1.43</td>
<td>-1.43</td>
</tr>
<tr>
<td></td>
<td>-1.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>QS</strong></td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Loss is incurred** (in a finite sample) by being approximate. But it is negligible. Computational gain? 

\[ p_{\text{exact}}(y_{T+1} | y_T) : 360 \text{ seconds} \]

\[ p_{\text{ABC}}(y_{T+1} | y_T) : 3 \text{ seconds!} \]

(with parallel computing)

Gael Martin, Monash University, Melbourne, Forecasting with Approximate Bayesian Comp
Expected scores: exact and ABC: $T = 500$

- **Average predictive scores** over 500 out-of-sample values:

<table>
<thead>
<tr>
<th></th>
<th>$\eta^{(1)}(y)$</th>
<th>$\eta^{(2)}(y)$</th>
<th>$\eta^{(3)}(y)$</th>
<th>$\eta^{(4)}(y)$</th>
<th>Exact av. score</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS</td>
<td>-1.43</td>
<td>-1.42</td>
<td>-1.43</td>
<td>-1.43</td>
<td>-1.40</td>
</tr>
<tr>
<td>QS</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.29</td>
</tr>
<tr>
<td>CRPS</td>
<td>-0.57</td>
<td><strong>-0.56</strong></td>
<td>-0.57</td>
<td>-0.57</td>
<td>-0.56</td>
</tr>
</tbody>
</table>

Loss is incurred (in a finite sample) by being approximate. But it is negligible.

Computational gain?

$p^{\text{exact}}(y_{T+1} | y_1) = 360$ seconds

$p^{\text{ABC}}(y_{T+1} | y_1) = 3$ seconds! (with parallel computing)
### Average predictive scores over 500 out-of-sample values:

<table>
<thead>
<tr>
<th></th>
<th>ABC av. score</th>
<th>Exact av. score</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta^{(1)}(y) )</td>
<td>-1.43</td>
<td>-1.40</td>
</tr>
<tr>
<td>( \eta^{(2)}(y) )</td>
<td>-1.42</td>
<td>-1.40</td>
</tr>
<tr>
<td>( \eta^{(3)}(y) )</td>
<td>-1.43</td>
<td>-1.40</td>
</tr>
<tr>
<td>( \eta^{(4)}(y) )</td>
<td>-1.43</td>
<td>-1.40</td>
</tr>
<tr>
<td>LS</td>
<td>-1.43</td>
<td>-1.40</td>
</tr>
<tr>
<td>QS</td>
<td>0.28</td>
<td>0.29</td>
</tr>
<tr>
<td>CRPS</td>
<td>-0.57</td>
<td>-0.56</td>
</tr>
<tr>
<td></td>
<td>-0.56</td>
<td>-0.56</td>
</tr>
</tbody>
</table>

- Loss is incurred (in a finite sample) by being approximate.
Expected scores: exact and ABC: $T = 500$

**Average predictive scores** over 500 out-of-sample values:

<table>
<thead>
<tr>
<th></th>
<th>$\eta^{(1)}(y)$</th>
<th>$\eta^{(2)}(y)$</th>
<th>$\eta^{(3)}(y)$</th>
<th>$\eta^{(4)}(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LS</strong></td>
<td>-1.43</td>
<td>-1.42</td>
<td>-1.43</td>
<td>-1.43</td>
</tr>
<tr>
<td><strong>QS</strong></td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td><strong>CRPS</strong></td>
<td>-0.57</td>
<td>-0.56</td>
<td>-0.57</td>
<td>-0.57</td>
</tr>
</tbody>
</table>

**Loss** is incurred *(in a finite sample)* by being **approximate**

But it is **negligible**
Expected scores: exact and ABC: $T = 500$

- **Average predictive scores** over 500 out-of-sample values:

<table>
<thead>
<tr>
<th></th>
<th>$\eta^{(1)}(y)$</th>
<th>$\eta^{(2)}(y)$</th>
<th>$\eta^{(3)}(y)$</th>
<th>$\eta^{(4)}(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS</td>
<td>-1.43</td>
<td>-1.42</td>
<td>-1.43</td>
<td>-1.43</td>
</tr>
<tr>
<td>QS</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td>CRPS</td>
<td>-0.57</td>
<td>-0.56</td>
<td>-0.57</td>
<td>-0.57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Exact av. score</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS</td>
<td>-1.40</td>
</tr>
<tr>
<td>QS</td>
<td>0.29</td>
</tr>
<tr>
<td>CRPS</td>
<td>-0.56</td>
</tr>
</tbody>
</table>

- Loss is incurred (**in a finite sample**) by being **approximate**
- But it is **negligible**
- Computational gain?
Expected scores: exact and ABC: $T = 500$

- Average predictive scores over 500 out-of-sample values:

<table>
<thead>
<tr>
<th></th>
<th>$\eta^{(1)}(y)$</th>
<th>$\eta^{(2)}(y)$</th>
<th>$\eta^{(3)}(y)$</th>
<th>$\eta^{(4)}(y)$</th>
<th>Exact av. score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LS</strong></td>
<td>-1.43</td>
<td>-1.42</td>
<td>-1.43</td>
<td>-1.43</td>
<td>-1.40</td>
</tr>
<tr>
<td><strong>QS</strong></td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.29</td>
</tr>
<tr>
<td><strong>CRPS</strong></td>
<td>-0.57</td>
<td>-0.56</td>
<td>-0.57</td>
<td>-0.57</td>
<td>-0.56</td>
</tr>
</tbody>
</table>

- Loss is incurred (in a finite sample) by being approximate
- But it is negligible
- Computational gain?
  - $p_{\text{exact}}(y_{T+1}|y)$: 360 seconds
Expected scores: exact and ABC: $T = 500$

- **Average predictive scores** over 500 out-of-sample values:

<table>
<thead>
<tr>
<th></th>
<th>$\eta^{(1)}(y)$</th>
<th>$\eta^{(2)}(y)$</th>
<th>$\eta^{(3)}(y)$</th>
<th>$\eta^{(4)}(y)$</th>
<th>Exact av. score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LS</strong></td>
<td>-1.43</td>
<td>-1.42</td>
<td>-1.43</td>
<td>-1.43</td>
<td>-1.40</td>
</tr>
<tr>
<td><strong>QS</strong></td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.29</td>
</tr>
<tr>
<td><strong>CRPS</strong></td>
<td>-0.57</td>
<td>-0.56</td>
<td>-0.57</td>
<td>-0.57</td>
<td>-0.56</td>
</tr>
</tbody>
</table>

- Loss **is incurred** *(in a finite sample)* by being **approximate**
- But it is **negligible**
- Computational gain?
  - $p_{exact}(y_{T+1}|y)$: 360 seconds
  - $p_{ABC}(y_{T+1}|y)$: 3 seconds!
Expected scores: exact and ABC: \( T = 500 \)

- **Average predictive scores** over 500 out-of-sample values:

<table>
<thead>
<tr>
<th></th>
<th>( \eta^{(1)}(y) )</th>
<th>( \eta^{(2)}(y) )</th>
<th>( \eta^{(3)}(y) )</th>
<th>( \eta^{(4)}(y) )</th>
<th>Exact av. score</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS</td>
<td>-1.43</td>
<td>-1.42</td>
<td>-1.43</td>
<td>-1.43</td>
<td>-1.40</td>
</tr>
<tr>
<td>QS</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.29</td>
</tr>
<tr>
<td>CRPS</td>
<td>-0.57</td>
<td>-0.56</td>
<td>-0.57</td>
<td>-0.57</td>
<td>-0.56</td>
</tr>
</tbody>
</table>

- Loss **is incurred** (in a finite sample) by being **approximate**
- But it is **negligible**
- Computational gain?
  - \( p_{\text{exact}}(y_{T+1} | y) \): 360 seconds
  - \( p_{\text{ABC}}(y_{T+1} | y) \): 3 seconds!
**Expected scores: exact and ABC: \( T = 500 \)**

- **Average predictive scores** over 500 out-of-sample values:

<table>
<thead>
<tr>
<th></th>
<th>( \eta^{(1)}(y) )</th>
<th>( \eta^{(2)}(y) )</th>
<th>( \eta^{(3)}(y) )</th>
<th>( \eta^{(4)}(y) )</th>
<th>Exact av. score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LS</strong></td>
<td>-1.43</td>
<td>-1.42</td>
<td>-1.43</td>
<td>-1.43</td>
<td>-1.40</td>
</tr>
<tr>
<td><strong>QS</strong></td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.29</td>
</tr>
<tr>
<td><strong>CRPS</strong></td>
<td>-0.57</td>
<td>-0.56</td>
<td>-0.57</td>
<td>-0.57</td>
<td>-0.56</td>
</tr>
</tbody>
</table>

- Loss **is incurred** (**in a finite sample**) by being **approximate**
- But it is **negligible**
- Computational gain?
  - \( p_{\text{exact}}(y_{T+1}|y) \) : 360 seconds
  - \( p_{\text{ABC}}(y_{T+1}|y) \) : 3 seconds! (with parallel computing)
ABC prediction in state space models?

- How does one compute $p_{ABC}(y_{T+1} | y)$ in state space models?
ABC prediction in state space models?

- How does one compute $p_{ABC}(y_{T+1}|y)$ in state space models?
  - Does one condition state inference only on $\eta(y)$?
ABC prediction in state space models?

- How does one compute $p_{ABC}(y_{T+1}|y)$ in state space models?
  - Does one condition state inference only on $\eta(y)$?

- Given a financial return, $y_t = \ln P_t - \ln P_{t-1}$
ABC prediction in state space models?

- How does one compute $p_{ABC}(y_{T+1}|y)$ in state space models?
- Does one condition state inference only on $\eta(y)$?

Given a financial return, $y_t = \ln P_t - \ln P_{t-1}$

Assume stochastic volatility:

$$y_t = \sqrt{V_t}\varepsilon_t; \quad \varepsilon_t \sim i.i.d. N(0, 1)$$
ABC prediction in state space models?

- How does one compute $p_{ABC}(y_{T+1}|y)$ in state space models?
  - Does one condition state inference only on $\eta(y)$?
- Given a financial return, $y_t = \ln P_t - \ln P_{t-1}$
- Assume stochastic volatility:
  \[ y_t = \sqrt{V_t}\varepsilon_t; \quad \varepsilon_t \sim i.i.d. N(0, 1) \]
How does one compute $p_{ABC}(y_{T+1}|y)$ in state space models?

Does one condition state inference only on $\eta(y)$?

Given a financial return, $y_t = \ln P_t - \ln P_{t-1}$

Assume stochastic volatility:

$$y_t = \sqrt{V_t} \varepsilon_t; \quad \varepsilon_t \sim i.i.d. N(0, 1)$$

$$\ln V_t = \theta_1 \ln V_{t-1} + \eta_t; \quad \eta_t \sim i.i.d. N(0, \theta_2)$$
ABC prediction in state space models?

- How does one compute $p_{ABC}(y_{T+1}|y)$ in state space models?
  - Does one condition state inference only on $\eta(y)$?
- Given a financial return, $y_t = \ln P_t - \ln P_{t-1}$
- Assume stochastic volatility:
  
  $$y_t = \sqrt{V_t}\varepsilon_t; \quad \varepsilon_t \sim i.i.d. N(0, 1)$$

  $$\ln V_t = \theta_1 \ln V_{t-1} + \eta_t; \quad \eta_t \sim i.i.d. N(0, \theta_2)$$

  $$\theta = (\theta_1, \theta_2)'$$
ABC prediction in state space models?

- **Exact:**

$$p_{\text{exact}}(y_{T+1} | y) = \int_{V_{T+1}} \int_{\mathbf{V}} \int_{\theta} p(y_{T+1} | V_{T+1})$$

$$\times p(V_{T+1} | V_T, y, \theta) p(\mathbf{V} | \theta, y) p(\theta | y) d\theta d\mathbf{V} dV_{T+1}$$

Gael Martin, Monash University, Melbourne. Forecasting with Approximate Bayesian Comp
ABC prediction in state space models?

- **Exact:**

\[
p_{\text{exact}}(y_{T+1}|y) = \int_{V_{T+1}} \int_{\mathbf{V}} \int_{\theta} p(y_{T+1}|V_{T+1}) \times p(V_{T+1}|V_T, y, \theta) p(V|\theta, y) p(\theta|y) d\theta d\mathbf{V} dV_{T+1} \]

- **MCMC** used to draw from \( p(V, \theta|y) \)
ABC prediction in state space models?

**Exact:**

\[
p_{\text{exact}}(y_{T+1}|y) = \int_{V_{T+1}} \int_{\theta} \int_{V} p(y_{T+1}|V_{T+1}) \\
\times p(V_{T+1}|V_T, y, \theta) p(V|\theta, y) p(\theta|y) d\theta dV dV_{T+1} \\
p(V, \theta|y)
\]

- **MCMC** used to draw from \( p(V, \theta|y) \)
- \( \Rightarrow \text{independent} \) draws from \( p(V_{T+1}|V_T, y, \theta) \) and \( p(y_{T+1}|V_{T+1}) \) \( \Rightarrow \) \( y_{T+1}^{(i)} \)
ABC prediction in state space models?

- **Exact:**

\[
p_{\text{exact}}(y_{T+1}|y) = \int_{V_{T+1}} \int_{\theta} \int_{V} p(y_{T+1}|V_{T+1}) \times p(V_{T+1}|V_T, y, \theta) p(V|\theta, y) p(\theta|y) d\theta dV dV_{T+1} \]

- **MCMC** used to draw from \( p(V, \theta|y) \)

\[ \Rightarrow \text{independent} \text{ draws from } p(V_{T+1}|V_T, y, \theta) \text{ and } \]
\[ p(y_{T+1}|V_{T+1}) \Rightarrow y_{T+1}^{(i)} \]

\[ \Rightarrow \overset{\text{p}}{p}_{\text{exact}}(y_{T+1}|y) \]
ABC prediction in state space models?

ABC:

\[
p_{ABC}(y_{T+1}|y) = \int_{V_{T+1}} \int_{V} \int_{\theta} p(y_{T+1}|V_{T+1}) \\
\times p(V_{T+1}|V, y, \theta) p(V|\theta, y) p(\theta|\eta(y)) d\theta dV dV_{T+1}
\]
ABC prediction in state space models?

- **ABC:**

\[
p_{ABC}(y_{T+1}|y) = \int_{V_{T+1}} \int_{V} \int_{\theta} p(y_{T+1}|V_{T+1}) \\
\times p(V_{T+1}|V, y, \theta) p(V|\theta, y) p(\theta|\eta(y)) \, d\theta \, dV \, dV_{T+1}
\]

- **ABC** used to draw from \( p(\theta|\eta(y)) \)
ABC prediction in state space models?

- **ABC:**

\[
p_{ABC}(y_{T+1}|y) = \int_{V_{T+1}} \int_{V} \int_{\theta} p(y_{T+1}|V_{T+1}) \times p(V_{T+1}|V_{T}, y, \theta)p(V|\theta, y)p(\theta|\eta(y))d\theta dV dV_{T+1}
\]

- **ABC** used to draw from \(p(\theta|\eta(y))\)

- (with \(\eta(y)\) based on an approximating auxiliary GARCH model)
ABC prediction in state space models?

- **ABC:**

\[
p_{ABC}(y_{T+1}|y) = \int_{V_{T+1}} \int_{V} \int_{\theta} p(y_{T+1}|V_{T+1}) \\
\times p(V_{T+1}|V_{T}, y, \theta)p(V|\theta, y)p(\theta|\eta(y))d\theta dV dV_{T+1}
\]

- ABC used to draw from \(p(\theta|\eta(y))\)
- (with \(\eta(y)\) based on an approximating auxiliary GARCH model)
- \(\Rightarrow\) particle filtering used to integrate out \(V\)
ABC prediction in state space models?

- ABC:

\[
p_{ABC}(y_{T+1}|y) = \int_{V_{T+1}} \int_{\theta} \int_{V_T} p(y_{T+1}|V_{T+1}) \\
\times p(V_{T+1}|V_T, y, \theta) p(V|\theta, y) p(\theta|\eta(y)) \, d\theta \, dV \, dV_{T+1}
\]

- ABC used to draw from \( p(\theta|\eta(y)) \)
- (with \( \eta(y) \) based on an approximating auxiliary GARCH model)
- \( \Rightarrow \) particle filtering used to integrate out \( V \)
- \( \Rightarrow \) yields full posterior inference (i.e. \( |y \)) on \( V_T \)
ABC prediction in state space models?

- **ABC:**

\[
p_{ABC}(y_{T+1}|y) = \int_{V_{T+1}} \int_{V} \int_{\theta} p(y_{T+1}|V_{T+1})
\times p(V_{T+1}|V, y, \theta)p(V|\theta, y)p(\theta|\eta(y))d\theta dV dV_{T+1}
\]

- **ABC** used to draw from \(p(\theta|\eta(y))\)

  - (with \(\eta(y)\) based on an approximating **auxiliary** GARCH model)

  \(\Rightarrow\) **particle filtering** used to integrate out \(V\)

  \(\Rightarrow\) yields **full posterior inference** (i.e. \(|y|\) on \(V_T\))

  - Exact inference (MCMC) on \(V_{1:T-1}\) not required
Nature of ABC inference on $\theta$ of little importance.

$p_{\text{ABC}}(y_T + 1 | \eta(y))$

What if condition $V_T$ on $\eta(y)$ only? i.e. omit the PF step?

The green curve (i.e. inaccuracy!)

Gael Martin, Monash University, Melbourne. Forecasting with Approximate Bayesian Comp

18 / 35
Nature of ABC inference on $\theta$ of little importance...

$p_{\text{ABC}}(y_{T+1} | y_T + 1)$

What if condition $V_T$ on $\eta(y)$ only? i.e. omit the PF step?

Gael Martin, Monash University, Melbourne, IMS Workshop, August, 2018
- Nature of ABC inference on $\theta$ of little importance.....
Nature of ABC inference on $\theta$ of little importance.....

$\Rightarrow \textbf{all } p_{ABC}(y_{T+1} | y) \approx p_{exact}(y_{T+1} | y)!$
Nature of ABC inference on $\theta$ of little importance...

$\Rightarrow \textbf{all } p_{ABC}(y_{T+1}|y) \approx p_{\text{exact}}(y_{T+1}|y)!$

What if condition $V_T$ on $\eta(y)$ only? i.e. omit the PF step?
- Nature of ABC inference on $\theta$ of little importance.....
- $\Rightarrow \text{all } p_{ABC}(y_{T+1}|y) \approx p_{exact}(y_{T+1}|y)!$
- What if condition $V_T$ on $\eta(y)$ only? i.e. omit the PF step?
- $\Rightarrow$ the green curve (i.e. inaccuracy!)
ABC prediction in state space models?

Need to get the predictive model: \( p(y_{T+1} \mid V_{T+1}) \) and \( p(V_{T+1} \mid V_T, y, \theta) \) right!
ABC prediction in state space models?

- Need to get the predictive model: $p(y_{T+1} | V_{T+1})$ and $p(V_{T+1} | V_T, y, \theta)$ right!
- But only need particle filtering to do that
ABC prediction in state space models?

- Need to get the predictive **model**: \( p(y_{T+1} | V_{T+1}) \) and \( p(V_{T+1} | V_T, y, \theta) \) right!

- But only need **particle filtering** to do that

\[ \Rightarrow \text{ABC prediction still based on independent sampling} \]
ABC prediction in state space models?

- Need to get the predictive model: \( p(y_{T+1} | V_{T+1}) \) and \( p(V_{T+1} | V_T, y, \theta) \) right!

- But only need particle filtering to do that

\[ \Rightarrow \text{ABC prediction still based on independent sampling} \]

\[ \Rightarrow \text{parallel computing can still be exploited} \]
Empirical setting?

- Thus far?
Empirical setting??

Thus far?

1. That the DGP: $p(y_{T+1}, y_j \theta) = p(y_{T+1} \theta \eta(y)) p(y_j \theta)$ is correctly specified (whether latent states are playing a role or not....)

2. That we have access to $p(\theta_j y)$ $p_{exact}(y_{T+1} \eta(y))$ for assessment of $p(\theta_j \eta(y)) p_{ABC}(y_{T+1} \eta(y))$

In a realistic empirical setting:

1. The assumed DGP will be misspecified

2. We are accessing $p_{ABC}(y_{T+1} \eta(y))$ because we cannot (or it is too computationally burdensome) to access $p_{exact}(y_{T+1} \eta(y))$ no benchmark for $p_{ABC}(y_{T+1} \eta(y))$

3. Critically.....the sense in which $p_{exact}(y_{T+1} \eta(y))$ remains the gold standard is no longer clear

Gael Martin, Monash University, Melbourne. Forecasting with Approximate Bayesian Comp
Empirical setting??

- Thus far? Have assumed:

\[ p(y_{T+1}, y_{j|\theta}) = p(y_{j|\theta}) p(y_{T+1} | y_{j|\theta}) \]

In a realistic empirical setting:

1. The assumed DGP will be misspecified.
2. We are accessing \( p_{ABC}(y_{T+1} | y_{j|\theta}) \) because we cannot (or it is too computationally burdensome) to access \( p_{exact}(y_{T+1} | y_{j|\theta}) \).

Critically, the sense in which \( p_{exact}(y_{T+1} | y_{j|\theta}) \) remains the gold standard is no longer clear.
Empirical setting??

Thus far? Have assumed:

1. That the **DGP**: \(p(y_{T+1}, y | \theta) = p(y_{T+1} | y, \theta) p(y | \theta)\) is **correctly specified**

Gael Martin, Monash University, Melbourne. Forecasting with Approximate Bayesian Comp
Thus far? Have assumed:

1. That the **DGP**: \( p(y_{T+1}, y|\theta) = p(y_{T+1}|y, \theta)p(y|\theta) \) is correctly specified
   - (whether latent states are playing a role or not.....)
Empirical setting??

Thus far? Have assumed:

1. That the DGP: \( p(y_{T+1}, y|\theta) = p(y_{T+1}|y, \theta)p(y|\theta) \) is correctly specified
   (whether latent states are playing a role or not.....)

2. That we have access to \( p(\theta|y) \Rightarrow p_{\text{exact}}(y_{T+1}|y) \)
Empirical setting??

Thus far? Have assumed:

1. That the DGP: $p(y_{T+1}, \theta | y) = p(y_{T+1} | y, \theta)p(y | \theta)$ is **correctly specified**
   - (whether latent states are playing a role or not.....)

2. That we have access to $p(\theta | y) \Rightarrow p_{\text{exact}}(y_{T+1} | y)$
   - for assessment of $p(\theta | \eta(y)) \Rightarrow p_{ABC}(y_{T+1} | y)$
Empirical setting??

Thus far? Have assumed:

1. That the **DGP**: \( p(y_{T+1}, y|\theta) = p(y_{T+1}|y, \theta)p(y|\theta) \) is **correctly specified**
   - (whether latent states are playing a role or not.....)

2. That we have access to \( p(\theta|y) \Rightarrow p_{\text{exact}}(y_{T+1}|y) \)
   - for assessment of \( p(\theta|\eta(y)) \Rightarrow p_{ABC}(y_{T+1}|y) \)

In a realistic empirical setting:
Thus far? Have assumed:

1. That the **DGP:** \( p(y_{T+1}, y | \theta) = p(y_{T+1} | y, \theta) p(y | \theta) \) is **correctly specified**
   - (whether latent states are playing a role or not.....)

2. That we have access to \( p(\theta | y) \Rightarrow p_{\text{exact}}(y_{T+1} | y) \)
   - for assessment of \( p(\theta | \eta(y)) \Rightarrow p_{\text{ABC}}(y_{T+1} | y) \)

In a realistic empirical setting:

1. The assumed **DGP** will be **misspecified**
Empirical setting??

- Thus far? Have assumed:

1. That the DGP: \( p(y_{T+1}, y | \theta) = p(y_{T+1} | y, \theta)p(y | \theta) \) is correctly specified
   - (whether latent states are playing a role or not.....)

2. That we have access to \( p(\theta | y) \Rightarrow p_{exact}(y_{T+1} | y) \)
   - for assessment of \( p(\theta | \eta(y)) \Rightarrow p_{ABC}(y_{T+1} | y) \)

- In a realistic empirical setting:

1. The assumed DGP will be misspecified
2. We are accessing \( p_{ABC}(y_{T+1} | y) \) because we cannot (or it is too computationally burdensome) to access \( p_{exact}(y_{T+1} | y) \)
Empirical setting??

Thus far? Have assumed:

1. That the DGP: \( p(y_{T+1}, y|\theta) = p(y_{T+1}|y, \theta)p(y|\theta) \) is correctly specified
   - (whether latent states are playing a role or not.....)

2. That we have access to \( p(\theta|y) \Rightarrow p_{\text{exact}}(y_{T+1}|y) \)
   - for assessment of \( p(\theta|\eta(y)) \Rightarrow p_{\text{ABC}}(y_{T+1}|y) \)

In a realistic empirical setting:

1. The assumed DGP will be misspecified
2. We are accessing \( p_{\text{ABC}}(y_{T+1}|y) \) because we cannot (or it is too computationally burdensome) to access \( p_{\text{exact}}(y_{T+1}|y) \)
   - \( \Rightarrow \) no benchmark for \( p_{\text{ABC}}(y_{T+1}|y) \)
Empirical setting??

Thus far? Have assumed:

1. That the DGP: \( p(y_{T+1}, y | \theta) = p(y_{T+1} | y, \theta)p(y | \theta) \) is correctly specified
   - (whether latent states are playing a role or not.....)

2. That we have access to \( p(\theta | y) \Rightarrow p_{exact}(y_{T+1} | y) \)
   - for assessment of \( p(\theta | \eta(y)) \Rightarrow p_{ABC}(y_{T+1} | y) \)

In a realistic empirical setting:

1. The assumed DGP will be misspecified
2. We are accessing \( p_{ABC}(y_{T+1} | y) \) because we cannot (or it is too computationally burdensome) to access \( p_{exact}(y_{T+1} | y) \)
   - \( \Rightarrow \) no benchmark for \( p_{ABC}(y_{T+1} | y) \)
3. Critically.....the sense in which \( p_{exact}(y_{T+1} | y) \) remains the gold standard is no longer clear
Two routes:
Empirical setting??

- Two routes:
  1. Choose a range of different $\eta(y)$ (and, hence, $p(\theta|\eta(y)))$
Empirical setting??

Two routes:

1. Choose a range of different $\eta(y)$ (and, hence, $p(\theta|\eta(y)))$ → a range of different $p_{ABC}(y_{T+1}|y)$
Empirical setting??

Two routes:

1. Choose a range of different $\eta(y)$ (and, hence, $p(\theta|\eta(y))$)
   - $\Rightarrow$ a range of different $p_{ABC}(y_{T+1}|y)$
   - select that $p_{ABC}(y_{T+1}|y)$ (and hence $p(\theta|\eta(y))$) according to **predictive performance** in a hold-out sample
Empirical setting??

Two routes:

1. Choose a range of different $\eta(y)$ (and, hence, $p(\theta|\eta(y)))$
   - $\Rightarrow$ a range of different $p_{ABC}(y_{T+1}|y)$
   - Select that $p_{ABC}(y_{T+1}|y)$ (and hence $p(\theta|\eta(y)))$ according to predictive performance in a hold-out sample
   - (‘ABF’, 2018)
Empirical setting??

Two routes:

1. Choose a range of different $\eta(y)$ (and, hence, $p(\theta|\eta(y)))$

   - $\Rightarrow$ a range of different $p_{ABC}(y_{T+1}|y)$
   - select that $p_{ABC}(y_{T+1}|y)$ (and hence $p(\theta|\eta(y)))$ according to predictive performance in a hold-out sample

   - (‘ABF’, 2018)

   - $\eta(y)$ still chosen to be informative about $\theta$
Empirical setting??

Two routes:

1. Choose a range of different $\eta(y)$ (and, hence, $p(\theta | \eta(y))$)
   - $\Rightarrow$ a range of different $p_{ABC}(y_{T+1} | y)$
   - select that $p_{ABC}(y_{T+1} | y)$ (and hence $p(\theta | \eta(y))$) according to predictive performance in a hold-out sample
   - ('ABF', 2018)
   - $\eta(y)$ still chosen to be informative about $\theta$

2. Choose $\eta(y)$ according to a predictive criterion
Two routes:

1. Choose a range of different $\eta(y)$ (and, hence, $p(\theta|\eta(y))$)
   - $\Rightarrow$ a range of different $\rho_{ABC}(y_{T+1}|y)$
   - select that $\rho_{ABC}(y_{T+1}|y)$ (and hence $p(\theta|\eta(y))$) according to predictive performance in a hold-out sample
   - (‘ABF’, 2018)
   - $\eta(y)$ still chosen to be informative about $\theta$

2. Choose $\eta(y)$ according to a predictive criterion
   - $\Rightarrow \eta(y) = f^n(S(p, y_{T+1}))$
Empirical setting??

Two routes:

1. Choose a range of different $\eta(y)$ (and, hence, $p(\theta|\eta(y))$)
   - $\Rightarrow$ a range of different $p_{ABC}(y_{T+1}|y)$
   - select that $p_{ABC}(y_{T+1}|y)$ (and hence $p(\theta|\eta(y))$) according to predictive performance in a hold-out sample
   - (‘ABF’, 2018)
   - $\eta(y)$ still chosen to be informative about $\theta$

2. Choose $\eta(y)$ according to a predictive criterion
   - $\Rightarrow \eta(y) = f^n(S(p, y_{T+1}))$
   - How to choose $S$?
Empirical setting??

Two routes:

1. Choose a range of different $\eta(y)$ (and, hence, $p(\theta|\eta(y))$)
   
   - $\Rightarrow$ a range of different $p_{ABC}(y_{T+1}|y)$
   - select that $p_{ABC}(y_{T+1}|y)$ (and hence $p(\theta|\eta(y))$) according to predictive performance in a hold-out sample
   - (‘ABF’, 2018)
   - $\eta(y)$ still chosen to be informative about $\theta$

2. Choose $\eta(y)$ according to a predictive criterion
   
   - $\Rightarrow \eta(y) = f^n(S(p, y_{T+1}))$
   - How to choose $S$?
   - How to specify $f^n(.)$?
Empirical setting??

- Two routes:
  1. Choose a range of different $\eta(y)$ (and, hence, $p(\theta|\eta(y))$)
     - $\Rightarrow$ a range of different $p_{ABC}(y_{T+1}|y)$
     - select that $p_{ABC}(y_{T+1}|y)$ (and hence $p(\theta|\eta(y))$) according to **predictive performance** in a hold-out sample
     - (‘ABF’, 2018)
     - $\eta(y)$ still chosen to be informative about $\theta$
  2. Choose $\eta(y)$ according to a **predictive criterion**
     - $\Rightarrow \eta(y) = f^n(S(p, y_{T+1}))$
     - How to choose $S$?
     - How to specify $f^n(.)$?
     - How to **assess** the resulting approximate predictives?
Example of route 2

- True DGP (for log of asset price, $p_t = \ln P_t$):
Example of route 2

- **True DGP** (for log of asset price, $p_t = \ln P_t$):

- Jump diffusion with (square root) stochastic volatility:

\[
dp_t = \sqrt{V_t} dB_t^p + \underbrace{Z_t dN_t}_{= g(\theta_{0,4}, \theta_{0,5}, \ldots)}
\]

\[
dV_t = (\theta_{0,1} - \theta_{0,2} V_t) dt + \theta_{0,3} \sqrt{V_t} dB_t^v
\]
Example of route 2

- **True DGP** (for log of asset price, $p_t = \ln P_t$):

  - Jump diffusion with (square root) stochastic volatility:
    
    $$
    dp_t = \sqrt{V_t} dB_t^p + \underbrace{Z_t dN_t}_{g(\theta_{0,4}, \theta_{0,5}, \ldots)}
    $$
    
    $$
    dV_t = (\theta_{0,1} - \theta_{0,2} V_t) \, dt + \theta_{0,3} \sqrt{V_t} dB_t^v
    $$

- $\theta_0 = (\theta_{0,1}, \theta_{0,2}, \theta_{0,3}, \ldots)' = \text{true parameter (vector)}$
Example of route 2

- **True DGP** (for log of asset price, $p_t = \ln P_t$):

- Jump diffusion with (square root) stochastic volatility:

  $$dp_t = \sqrt{V_t} dB^p_t + \underbrace{Z_t dN_t}_{= g(\theta_{0,4}, \theta_{0,5}, \ldots)}$$

  $$dV_t = (\theta_{0,1} - \theta_{0,2} V_t)\, dt + \theta_{0,3} \sqrt{V_t} dB^v_t$$

  - $\theta_0 = (\theta_{0,1}, \theta_{0,2}, \theta_{0,3}, \ldots)' = \text{true parameter} (\text{vector})$

- **Assume:**

  $$dp_t = \sqrt{V_t} dB^p_t$$

  $$dV_t = (\theta_1 - \theta_2 V_t)\, dt + \theta_3 \sqrt{V_t} dB^v_t$$
Example of route 2

\[ \Rightarrow \text{ implies a model for } y_t = \ln P_t - \ln P_{t-1} \text{ (return at time } t) : \]
Example of route 2

\[ y_t = \ln P_t - \ln P_{t-1} \] (return at time \( t \)):

- which is **mis-specified**
Example of route 2

\[ y_t = \ln P_t - \ln P_{t-1} \] (return at time \( t \)):

- which is mis-specified

\[ p(\theta|y) \] (under regularity) concentrates onto pseudo-true \( \theta, \theta^* \)
Example of route 2

- \( \Rightarrow \) implies a model for \( y_t = \ln P_t - \ln P_{t-1} \) (return at time \( t \)):

- which is **mis-specified**

- \( p(\theta|y) \) (under regularity) concentrates onto **pseudo-true** \( \theta, \theta^* \)

- where \( \theta^* \) is close to \( \theta_0 \) (in KL-based sense)
Example of route 2

\[ \lim_{T \to \infty} p_{\text{exact}}(y_{T+1}|y) = p(y_{T+1}|y, \theta^*) \]
Example of route 2

\[ \lim_{T \to \infty} p_{\text{exact}}(y_{T+1}|y) = p(y_{T+1}|y, \theta^*) \]
Example of route 2

\[
\lim_{T \to \infty} p_{\text{exact}}(y_{T+1} | \mathbf{y}) = p(y_{T+1} | \mathbf{y}, \theta^*) = \text{what??}
\]
Example of route 2

\[ \lim_{T \to \infty} p_{\text{exact}}(y_{T+1}|y) = p(y_{T+1}|y, \theta^*) = \text{what?} \]

- \( p \) is misspecified
Example of route 2

- \[ \lim_{T \to \infty} p_{\text{exact}}(y_{T+1} | y) = p(y_{T+1} | y, \theta^*) = \text{what??} \]

- \( p \) is misspecified

- \( \theta^* \neq \theta_0 \)
Example of route 2

\[ \lim_{T \to \infty} p_{\text{exact}}(y_{T+1} | y) = p(y_{T+1} | y, \theta^*) = \text{what??} \]

- \( p \) is misspecified
- \( \theta^* \neq \theta_0 \)
- And we have nowhere else to go.....
Example of route 2

\[ \lim_{T \to \infty} p_{\text{exact}}(y_{T+1}|y) = p(y_{T+1}|y, \theta^*) = \text{what??} \]

- \( p \) is misspecified
- \( \theta^* \neq \theta_0 \)
- And we have nowhere else to go.....
- With an \textbf{ABC-type} approach we have more room to move.....
Apply ABC-type principles: Option 1

1. Simulate
   \[ i = 1, 2, \ldots, N, i. \]
   \[ \text{draws of } \theta_i \text{ from } p(\theta). \]

2. Produce:
   \[ p(y_{T+1} | y, \theta_i) \] (using particle filter)

3. For each \( \theta_i \), evaluate score at observed \( y_{0:T+1} \):
   \[ S(p(y_{T+1} | y, \theta_i), y_{0:T+1}) \]

4. Over \( n_e \) observations in an evaluation period, compute:
   \[ \eta_i = \frac{1}{n_e} \sum_{\tau=0}^{n_e} S(p(y_{T+1+\tau} | y, \theta_i), y_{0:T+1+\tau}) \]

5. Select \( \theta_i \) such that:
   \[ \eta_i > \text{the highest } (\alpha\%, \text{say}) \text{ quantile} \]
Apply ABC-type principles: Option 1

1. Simulate $i = 1, 2, \ldots, N$, i.i.d. draws of $\theta^i$ from $p(\theta)$
Apply ABC-type principles: Option 1

1. Simulate $i = 1, 2, \ldots, N$, i.i.d. draws of $\theta_i$ from $p(\theta)$

2. Produce:

$$p(y_{T+1} | y, \theta^i) \quad \text{(using particle filter)}$$
Apply ABC-type principles: Option 1

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

2. Produce:

$$p(y_{T+1}|y, \theta^i) \quad \text{(using particle filter)}$$

3. For each $\theta^i$, evaluate score at observed $y^0_{T+1}$:

$$S(p(y_{T+1}|y, \theta^i), y^0_{T+1})$$
Apply ABC-type principles: Option 1

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

2. Produce:

   $$p(y_{T+1}|y, \theta^i) \text{ (using particle filter)}$$

3. For each $\theta^i$, evaluate score at observed $y_{T+1}^0$:

   $$S(p(y_{T+1}|y, \theta^i), y_{T+1}^0)$$

4. Over $n_e$ observations in an evaluation period, compute:

   $$\eta^i(.) = \frac{1}{n_e} \sum_{\tau=0}^{n_e} S(p(y_{T+1+\tau}|y_1; T+\tau, \theta^i), y_{T+1+\tau}^0)$$
Apply ABC-type principles: Option 1

1. Simulate $i = 1, 2, \ldots, N$, i.i.d. draws of $\theta^i$ from $p(\theta)$
2. Produce:

\[ p(y_{T+1}|y, \theta^i) \quad \text{(using particle filter)} \]

3. **For each $\theta^i$, evaluate score at observed $y_{T+1}^0$**:

\[ S(p(y_{T+1}|y, \theta^i), y_{T+1}^0) \]

4. Over $n_e$ observations in an evaluation period, compute:

\[ \eta^i(.) = \frac{1}{n_e} \sum_{\tau=0}^{n_e} S(p(y_{T+1+\tau}|y_1; T+\tau, \theta^i), y_{T+1+\tau}^0) \]

5. Select $\theta^i$ such that:

\[ \eta^i(.) > \text{the highest (}\alpha\%\text{, say) quantile} \]
Apply ABC-type principles: Option 1

- produces a range of plausible $p(y_{T+1+n_e} | y_{1:T+n_e}, \theta^i)$
Apply ABC-type principles: Option 1

- produces a range of plausible $p(y_{T+1+n_e} \mid y_{1:T+n_e}, \theta^i)$
- that match the $y^0_{T+1}$ well in terms of $S(p, y^0_{T+1})$
Apply ABC-type principles: Option 1

- produces a range of plausible $p(y_{T+1+n_e} | y_{1:T+n_e}, \theta^i)$
- that match the $y_{T+1}^0$ well in terms of $S(p, y_{T+1}^0)$
- Can be used to provide a simulation-based estimate of:

$$p_{ABC}(y_{T+1+n_e} | y_{1:T+n_e}) = \int p(y_{T+1+n_e} | y_{1:T+n_e}, \theta) p(\theta | \eta(.)) d\theta$$
Apply ABC-type principles: Option 1

- produces a range of **plausible** \( p(y_{T+1+n_e} | y_{1:T+n_e}, \theta^i) \)
- that **match** the \( y^0_{T+1} \) well in terms of \( S(p, y^0_{T+1}) \)
- Can be used to provide a simulation-based estimate of:

\[
p_{ABC}(y_{T+1+n_e} | y_{1:T+n_e}) = \int p(y_{T+1+n_e} | y_{1:T+n_e}, \theta) p(\theta | \eta(\cdot)) d\theta
\]

- By computing (over \( N_a \) ‘accepted’ \( \theta^i \)):

\[
p_{av}(y_{T+1+n_e} | y_{1:T+n_e}) = \frac{1}{N_a} \sum_{i=1}^{N_a} p(y_{T+1+n_e} | y_{1:T+n_e}, \theta^i)
\]
Apply ABC-type principles: Option 2

- Adopt the flavour of **auxiliary-model based** ABC
Apply ABC-type principles: Option 2

- Adopt the flavour of **auxiliary-model based** ABC
  - Drovandi et al., 2011, 2015, 2018; Creel and Kristensen, 2015; Drovandi, 2018
Apply ABC-type principles: Option 2

- Adopt the flavour of **auxiliary-model based** ABC
  - Drovandi et al., 2011, 2015, 2018; Creel and Kristensen, 2015; Drovandi, 2018
  - Martin, McCabe, Frazier, Maneesoonth. & Robert, 2018
Apply ABC-type principles: Option 2

- Adopt the flavour of **auxiliary-model based** ABC
  - Drovandi et al., 2011, 2015, 2018; Creel and Kristensen, 2015; Drovandi, 2018
  - Martin, McCabe, Frazier, Maneesoonth. & Robert, 2018
- Specify a **tractable** \( q(y_{T+1}, y, \beta) \) that approximates \( p(y_{T+1}, y, \theta) \)
Apply ABC-type principles: Option 2

- Adopt the flavour of **auxiliary-model based** ABC
  - Drovandi et al., 2011, 2015, 2018; Creel and Kristensen, 2015; Drovandi, 2018
  - Martin, McCabe, Frazier, Maneesoonth. & Robert, 2018

- Specify a **tractable** \( q(y_{T+1}, y, \beta) \) that approximates \( p(y_{T+1}, y, \theta) \)

- \( \hat{\beta}_{MLE} \Rightarrow \eta(y) \)
Apply ABC-type principles: Option 2

- Adopt the flavour of **auxiliary-model based** ABC
  - Drovandi et al., 2011, 2015, 2018; Creel and Kristensen, 2015; Drovandi, 2018
  - Martin, McCabe, Frazier, Maneesoonth. & Robert, 2018

- Specify a **tractable** $q(y_{T+1}, y, \beta)$ that approximates $p(y_{T+1}, y, \theta)$

- $\hat{\beta}_{MLE} \Rightarrow \eta(y)$

- Aim in auxiliary-model based ABC for **inference**?
Apply ABC-type principles: Option 2

- Adopt the flavour of **auxiliary-model based** ABC
  - Drovandi et al., 2011, 2015, 2018; Creel and Kristensen, 2015; Drovandi, 2018
  - Martin, McCabe, Frazier, Maneesoonth. & Robert, 2018

- Specify a **tractable** $q(y_{T+1}, y, \beta)$ that *approximates* $p(y_{T+1}, y, \theta)$

- $\hat{\beta}_{MLE} \Rightarrow \eta(y)$

- Aim in auxiliary-model based ABC for inference?

- Choose $q(y_{T+1}, y, \theta)$ to capture features of $p(y_{T+1}, y, \theta)$
Apply ABC-type principles: Option 2

- Adopt the flavour of **auxiliary-model based** ABC
  - Drovandi et al., 2011, 2015, 2018; Creel and Kristensen, 2015; Drovandi, 2018
  - Martin, McCabe, Frazier, Maneesoonth. & Robert, 2018

- Specify a **tractable** $q(y_{T+1}, y, \beta)$ that *approximates* $p(y_{T+1}, y, \theta)$

- $\hat{\beta}_{\text{MLE}} \Rightarrow \eta(y)$

- Aim in auxiliary-model based ABC for **inference**?
  - Choose $q(y_{T+1}, y, \theta)$ to capture features of $p(y_{T+1}, y, \theta)$
  - If $q(y_{T+1}, y, \theta)$ ‘nests’ (a **correctly specified**) $p(y_{T+1}, y, \theta)$
Apply ABC-type principles: Option 2

- Adopt the flavour of **auxiliary-model based** ABC
  - Drovandi et al., 2011, 2015, 2018; Creel and Kristensen, 2015; Drovandi, 2018
  - Martin, McCabe, Frazier, Maneesoonth. & Robert, 2018

- Specify a **tractable** \( q(y_{T+1}, y, \beta) \) that **approximates** \( p(y_{T+1}, y, \theta) \)

- \( \hat{\beta}_{MLE} \Rightarrow \eta(y) \)

- Aim in auxiliary-model based ABC for **inference**?

- Choose \( q(y_{T+1}, y, \theta) \) to capture features of \( p(y_{T+1}, y, \theta) \)

- If \( q(y_{T+1}, y, \theta) \) ‘nests’ (a **correctly specified**) \( p(y_{T+1}, y, \theta) \)
  - \( \Rightarrow \eta(y) = \hat{\beta}_{MLE} \) is **asymptotically sufficient** for \( \theta \)
Apply ABC-type principles: Option 2

- Adopt the flavour of **auxiliary-model based** ABC
  - Drovandi et al., 2011, 2015, 2018; Creel and Kristensen, 2015; Drovandi, 2018
  - Martin, McCabe, Frazier, Maneesoonth. & Robert, 2018
- Specify a **tractable** \( q(y_{T+1}, y, \beta) \) that **approximates** \( p(y_{T+1}, y, \theta) \)
- \( \hat{\beta}_{MLE} \Rightarrow \eta(y) \)
- Aim in auxiliary-model based ABC for **inference**?
- Choose \( q(y_{T+1}, y, \theta) \) to capture features of \( p(y_{T+1}, y, \theta) \)
- If \( q(y_{T+1}, y, \theta) \) ‘nests’ (a **correctly specified** \( p(y_{T+1}, y, \theta) \))
  - \( \Rightarrow \eta(y) = \hat{\beta}_{MLE} \) is **asymptotically sufficient** for \( \theta \)
  - \( \Rightarrow p(\theta|\eta(y)) = p(\theta|y) \) (for large \( T \))
Apply ABC-type principles: Option 2

- Adopt the flavour of **auxiliary-model based** ABC
  - Drovandi et al., 2011, 2015, 2018; Creel and Kristensen, 2015; Drovandi, 2018
  - Martin, McCabe, Frazier, Maneesoonth. & Robert, 2018
- Specify a **tractable** \( q(y_{T+1}, y, \beta) \) that approximates \( p(y_{T+1}, y, \theta) \)
- \( \hat{\beta}_{MLE} \Rightarrow \eta(y) \)
- Aim in auxiliary-model based ABC for **inference**?
  - Choose \( q(y_{T+1}, y, \theta) \) to capture features of \( p(y_{T+1}, y, \theta) \)
  - If \( q(y_{T+1}, y, \theta) \) ‘nests’ (a **correctly specified**) \( p(y_{T+1}, y, \theta) \)
    - \( \Rightarrow \eta(y) = \hat{\beta}_{MLE} \) is **asymptotically sufficient** for \( \theta \)
    - \( \Rightarrow p(\theta|\eta(y)) = p(\theta|y) \) (for large \( T \))
    - \( \Rightarrow \) ‘ideal’ \( q(y_{T+1}, y, \theta) \) is **highly parameterized**
Apply ABC-type principles: Option 2

- But that do we know about forecasting??
Apply ABC-type principles: Option 2

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

Simple parsimoneous models often forecast better than complex, highly parameterized (but incorrect) models....

⇒ Approach in auxiliary-model based ABC for forecasting?
Apply ABC-type principles: Option 2

- But that do we know about **forecasting**??
- **Simple parsimoneous** models often forecast better than complex, highly parameterized (but incorrect) models....
- ⇒ Approach in auxiliary-model based ABC for **forecasting**?
- Pick a **simple parsimoneous** ‘auxiliary predictive’:

\[ q(y_{T+1}|y_{1:T}, \beta) \]
Apply ABC-type principles: Option 2

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

- **Simple parsimoneous** models often forecast better than complex, highly parameterized (but incorrect) models....

- ⇒ Approach in auxiliary-model based ABC for **forecasting**?

- Pick a **simple parsimoneous ‘auxiliary predictive’**:

  \[ q(y_{T+1} | y_1:T, \beta) \]

- And select \( \theta^i \) (and, hence, \( p(y_{T+1} | y, \theta^i) ) \)
Apply ABC-type principles: Option 2

- But that do we know about forecasting??

- **Simple parsimoneous** models often forecast better than complex, highly parameterized (but incorrect) models....

- ⇒ Approach in auxiliary-model based ABC for forecasting?

- Pick a **simple parsimoneous ‘auxiliary predictive’**:

  \[ q(y_{T+1}|y_{1:T}, \beta) \]

- And select \( \theta^i \) (and, hence, \( p(y_{T+1}|y, \theta^i) \))

- such that the predictive performance of \( p(y_{T+1}|y, \theta^i) \) matches that of \( q(y_{T+1}|y_{1:T}, \beta) \)
Apply ABC-type principles: Option 2

- But that do we know about forecasting??

- **Simple parsimoneous** models often forecast better than complex, highly parameterized (but incorrect) models..

- ⇒ Approach in auxiliary-model based ABC for forecasting?

- Pick a simple parsimoneous ‘auxiliary predictive’:

\[ q(y_{T+1}|y_1:T, \beta) \]

- And select \( \theta^i \) (and, hence, \( p(y_{T+1}|y, \theta^i) \))

- such that the predictive performance of \( p(y_{T+1}|y, \theta^i) \) matches that of \( q(y_{T+1}|y_1:T, \beta) \)

- ⇒ Replace Steps 4. and 5. above with:
4. Over \( n_e \) observations in an evaluation period, compute:

\[
\eta^i(.) = \frac{1}{n_e} \sum_{\tau=0}^{n_e} \left| p(y_{T+1+\tau}^0|y_1:T+\tau, \theta^i) - q(y_{T+1+\tau}^0|y_1:T+\tau, \hat{\beta}) \right|
\]
4. Over $n_e$ observations in an evaluation period, compute:

$$\eta^i(.) = \frac{1}{n_e} \sum_{\tau=0}^{n_e} \left| p(y_{T+1+\tau}^0 \mid y_{1:T+\tau}, \theta^i) - q(y_{T+1+\tau}^0 \mid y_{1:T+\tau}, \hat{\beta}) \right|$$

5. Select $\theta^i$ such that:

$$\eta^i(.) < \text{the lowest } (\alpha\%, \text{ say}) \text{ quantile}$$
4. Over $n_e$ observations in an evaluation period, compute:

$$\eta^i(.) = \frac{1}{n_e} \sum_{\tau=0}^{n_e} \left| p(y_{T+1+\tau}^0 | y_{1:T+\tau}, \theta^i) - q(y_{T+1+\tau}^0 | y_{1:T+\tau}, \hat{\beta}) \right|$$

5. Select $\theta^i$ such that:

$$\eta^i(.) < \text{the lowest } (\alpha\%, \text{say}) \text{ quantile}$$

- Produces a simulation-based estimate of a different:

$$p_{ABC}(y_{T+1+n_e} | y_{1:T+n_e}) = \int p(y_{T+1+n_e} | y_{1:T+n_e}, \theta) p(\theta | \eta(.) \right) d\theta$$
4. Over $n_e$ observations in an evaluation period, compute:

$$\eta^i(.) = \frac{1}{n_e} \sum_{\tau=0}^{n_e} \left| p(y^0_{T+1+\tau}|y_{1:T+\tau}, \theta^i) - q(y^0_{T+1+\tau}|y_{1:T+\tau}, \hat{\beta}) \right|$$

5. Select $\theta^i$ such that:

$$\eta^i(.) < \text{the lowest} \ (\alpha\%, \ \text{say}) \ \text{quantile}$$

- Produces a simulation-based estimate of a different:

$$p_{ABC}(y_{T+1+n_e}|y_{1:T+n_e}) = \int p(y_{T+1+n_e}|y_{1:T+n_e}, \theta) p(\theta|\eta(.)) \ d\theta$$

- in which $\eta(.)$ reflects a different measure of predictive performance
(Very!) preliminary results

Choose $q(y_T + 1: y_1)$ to be a generalized autoregressive conditionally heteroscedastic (GARCH) model with Student $t$ errors:

Work-horse of empirical finance

Often hard to beat in prediction of returns!

Display (for $T + 1 + n$)

Plots of accepted predictives (Options 1 and 2)

Averaged predictives (Options 1 and 2)

i.e. estimates of $p_{ABC}(y_T + 1 + n: y_1: T + n)$

Roll the whole process forward:

Compute log scores for 25 one-step-ahead predictions for both estimates of $p_{ABC}(y_T + 1 + n: y_1: T + n)$
Choose \( q(y_{T+1}|y_{1:T}, \beta) \) to be a generalized autoregressive conditionally heteroscedastic (GARCH) model with Student \( t \) errors:
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
(Very!) preliminary results

- 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

- Often hard to beat in prediction of returns!
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
- Often hard to beat in prediction of returns!
- Display (for \( T + 1 + n_e \))
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
- Often hard to beat in prediction of returns!
- Display (for $T + 1 + n_e$)
  1. Plots of accepted predictives (Options 1 and 2)
Choose \( q(y_{T+1}|y_1:T, \beta) \) to be a \textbf{generalized autoregressive conditionally heteroscedastic (GARCH)} model with Student \( t \) errors:

- Work-horse of empirical finance
- Often hard to beat in prediction of returns!

Display (for \( T + 1 + n_e \))

1. Plots of accepted predictives (Options 1 and 2)
2. Averaged predictives (Options 1 and 2)
(Very!) preliminary results

- 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
- Often hard to beat in prediction of returns!
- Display (for \( T + 1 + n_e \))
  1. Plots of accepted predictives (Options 1 and 2)
  2. Averaged predictives (Options 1 and 2)
    - i.e. estimates of \( p_{ABC}(y_{T+1+n_e}|y_{1:T+n_e}) \)
(Very!) preliminary results

- Choose \( q(y_{T+1} | y_1: T, \beta) \) to be a \textit{generalized autoregressive conditionally heteroscedastic (GARCH)} model with Student \( t \) errors:

- Work-horse of empirical finance
- Often hard to beat in prediction of returns!
- Display (for \( T + 1 + n_e \))
  1. Plots of accepted predictives (Options 1 and 2)
  2. Averaged predictives (Options 1 and 2)
     - i.e. estimates of \( p_{ABC}(y_{T+1+n_e} | y_1: T+n_e) \)
- Roll the whole process forward:
(Very!) preliminary results

- 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
- Often hard to beat in prediction of returns!
- Display (for \( T + 1 + n_e \))
  1. Plots of accepted predictives (Options 1 and 2)
  2. Averaged predictives (Options 1 and 2)
     - i.e. estimates of \( p_{ABC}(y_{T+1+n_e} | y_1:T+n_e) \)

- Roll the whole process forward:
  
  - Compute **log scores** for 25 one-step-ahead predictions for both estimates of \( p_{ABC}(y_{T+1+n_e} | y_1:T+n_e) \)
Plots of accepted conditional predictives
Plots of accepted conditional predictives

\[
\text{Draws from the posterior dist. of: } p(y_T + 1 | y_1: T + n, \theta) \\
\text{With uncertainty about } \theta \text{ conditioned on } \eta(.).
\]

Gael Martin, Monash University, Melbourne. Forecasting with Approximate Bayesian Comp
Draws from the posterior dist. of: $p(y_{T+1+n_e} | y_{1:T+n_e}, \theta)$
Draws from the posterior dist. of: \( p(y_{T+1+n_e} | y_{1:T+n_e}, \theta) \)

With uncertainty about \( \theta \) conditioned on \( \eta(\cdot) \)
Plots of accepted conditional predictives

- Draws from the posterior dist. of: $p(y_{T+1+n_e}|y_{1:T+n_e}, \theta)$
- With uncertainty about $\theta$ conditioned on $\eta(.)$
- Could extract distributions at the 5th and 95th percentiles
Averaged accepted predictives

- Or **average**: to produce estimates of $p_{ABC}(y_{T+1+n_e} | y_{1:T+n_e})$:

![Graph showing probability distribution](image)
Averaged accepted predictives

- Or average: to produce estimates of $p_{ABC}(y_{T+1+n_e} | y_{1:T+n_e})$:

![Graph showing probability distributions for different scenarios]

- Median scores (over 25 one-step-ahead periods):
  - Option 1: -0.262
  - Option 2: -0.114
Averaged accepted predictives

- Or **average**: to produce estimates of $p_{ABC}(y_{T+1+n_e} | y_{1:T+n_e})$:

-Median scores (over 25 one-step-ahead periods):
  - **Option 1**: -0.262; **Option 2**: -0.114
Change the auxiliary predictive?

\[ \text{Choose } q \left( y_{T+1} + \tau_j y_1: T+\tau_j, \beta \right) \text{ as GARCH with normal errors:} \]

Expected to be a poorer 'benchmark' (given the jumps in the true DGP):

Median scores:

Option 1: -0.262;
Option 2: -0.131

Still helps - but less so.

Gael Martin, Monash University, Melbourne, IMS Workshop, August, 2018

Forecasting with Approximate Bayesian Comp
Change the auxiliary predictive?

- Choose \( q(y_{T+1+\tau} \mid y_{1:T+\tau}, \beta) \) as GARCH with normal errors:
Change the auxiliary predictive?

- Choose $q(y_{T+1+\tau}|y_1:T+\tau, \beta)$ as GARCH with normal errors:
  - Expected to be a poorer ‘benchmark’ (given the jumps in the true DGP):
Change the auxiliary predictive?

- Choose \( q(y_{T+1+\tau} | y_{1:T+\tau}, \beta) \) as GARCH with normal errors:
  - Expected to be a poorer ‘benchmark’ (given the jumps in the true DGP):
Change the auxiliary predictive?

- Choose $q(y_{T+1+\tau}|y_{1:T+\tau}, \beta)$ as \textbf{GARCH} with \textbf{normal} errors:
  - Expected to be a poorer ‘benchmark’ (given the \textbf{jumps} in the true DGP):

![Graph showing probability density function]
Choose \( q(y_{T+1+\tau}|y_{1:T+\tau}, \beta) \) as GARCH with normal errors:

- Expected to be a poorer ‘benchmark’ (given the jumps in the true DGP):

\[
\text{Median scores: Option 1: -0.262; Option 2: -0.131}
\]
Change the auxiliary predictive?

- Choose \( q(y_{T+1+\tau} | y_{1:T+\tau}, \beta) \) as \textbf{GARCH} with \textbf{normal} errors:
  - Expected to be a poorer ‘benchmark’ (given the \textbf{jumps} in the true DGP):

\[
q(y_{T+1+\tau} | y_{1:T+\tau}, \beta) = GARCH \text{ with normal errors:}
\]

- Median scores: \textbf{Option 1}: -0.262; \textbf{Option 2}: -0.131
- Still helps - but less so
Comparison with forecasting performance with \textbf{exact} but \textbf{mis-specified} predictive:
Comparison with forecasting performance with **exact** but **mis-specified** predictive:

What would we expect?
To come.....

- Comparison with forecasting performance with **exact** but **mis-specified** predictive:
  - What would we expect?

  - Given that:
    \[
    \lim_{T \to \infty} p_{\text{exact}}(y_{T+1}|y) = p(y_{T+1}|y, \theta^*)
    \]
To come…..

- Comparison with forecasting performance with **exact** but **mis-specified** predictive:

- What would we expect?

- Given that:

\[
\lim_{T \to \infty} p_{\text{exact}}(y_{T+1}|y) = p(y_{T+1}|y, \theta^*)
\]

- where \( \theta^* \) minimizes the KL divergence of the assumed model from the **true DGP**
To come.....

- Comparison with forecasting performance with **exact** but **mis-specified** predictive:
  - What would we expect?
  - Given that:
    \[
    \lim_{T \to \infty} p_{exact}(y_{T+1}|y) = p(y_{T+1}|y, \theta^*)
    \]
    where $\theta^*$ minimizes the KL divergence of the assumed model from the **true** DGP
  - Will $p_{exact}(y_{T+1}|y)$ still ‘win’ in terms of **log score**?
Comparison with forecasting performance with \textbf{exact} but \textbf{mis-specified} predictive:

What would we expect?

Given that:

\[
\lim_{T \to \infty} p_{\text{exact}}(y_{T+1} | y) = p(y_{T+1} | y, \theta^*)
\]

where \(\theta^*\) minimizes the KL divergence of the assumed model from the \textbf{true DGP}

Will \(p_{\text{exact}}(y_{T+1} | y)\) still ‘win’ in terms of \textbf{log score}?

But \(p_{\text{ABC}}(y_{T+1} | y)\) ‘win’ in terms of \textbf{alternative performance criteria} (that have informed \(\eta(.)\))?
To come....

- If so
To come....

- If so
- $\Rightarrow$ Ideas may have relevance **beyond** usual ABC scenario
To come....

- If so

  ⇒ Ideas may have relevance **beyond** usual ABC scenario

  ⇒ May prompt some thinking about the use of different conditioning information in Bayesian forecasting **per se**
To come....

- If so
  - Ideas may have relevance **beyond** usual ABC scenario
  - May prompt some thinking about the use of different conditioning information in Bayesian forecasting **per se**
  - Including the use of $q$ as a **regularization** technique of sorts
To come....

- If so
  - Ideas may have relevance **beyond** usual ABC scenario
  - May prompt some thinking about the use of different conditioning information in Bayesian forecasting **per se**
  - Including the use of $q$ as a **regularization** technique of sorts
- Also:
To come....

- If so
- ⇒ Ideas may have relevance **beyond** usual ABC scenario
- ⇒ May prompt some thinking about the use of different conditioning information in Bayesian forecasting **per se**
- Including the use of $q$ as a **regularization** technique of sorts
- Also:
  - Can we produce asymptotic results in $n_e \ (\Rightarrow \eta(.)$ and $\alpha\%$?
To come....

- If so
- ⇒ Ideas may have relevance **beyond** usual ABC scenario
- ⇒ May prompt some thinking about the use of different conditioning information in Bayesian forecasting **per se**
- Including the use of $q$ as a **regularization** technique of sorts
- Also:
  - Can we produce asymptotic results in $n_e \ (\Rightarrow \eta(.) )$ and $\alpha\%$?
  - to mimic those in $T$ and $\varepsilon$ in:
To come....

- If so

  ⇒ Ideas may have relevance **beyond** usual ABC scenario

  ⇒ May prompt some thinking about the use of different conditioning information in Bayesian forecasting **per se**

- Including the use of $q$ as a **regularization** technique of sorts

- Also:
  - Can we produce asymptotic results in $n_e \ (⇒ \ η(.))$ and $α\%$?
  - to mimic those in $T$ and $ε$ in:

    - **Frazier, Martin, Robert and Rousseau, 2018**
To come....

- If so

  ⇒ Ideas may have relevance **beyond** usual ABC scenario

  ⇒ May prompt some thinking about the use of different conditioning information in Bayesian forecasting **per se**

- Including the use of $q$ as a **regularization** technique of sorts

- Also:
  
  - Can we produce asymptotic results in $n_e \ (\Rightarrow \eta(\cdot))$ and $\alpha\%$?
  
  - to mimic those in $T$ and $\varepsilon$ in:

    - **Frazier, Martin, Robert and Rousseau, 2018**

- · · · · · all in good time.....