Forecasting with Approximate Bayesian Computation (ABC)

Gael Martin

Monash University, Melbourne

IMS Workshop

August, 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:** *International Journal of Forecasting*
2. Under **misspecification** of the **DGP**:
   - **Frazier and Martin, 2018:**
   - Very preliminary!!

Exact Bayesian forecasting

- Distribution of interest is:

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

- **Exact (marginal)** predictive = expectation of the **conditional** predictive
- **Conditional** predictive reflects the **assumed DGP**
- (on which $p(\boldsymbol{\theta}|\mathbf{y})$ is also based)

Exact Bayesian forecasting

- Given M draws from $p(\boldsymbol{\theta}|\mathbf{y})$ (via a Markov chain Monte Carlo algorithm, say)
- $p_{exact}(y_{T+1}|\mathbf{y})$ can be **estimated** as
  1. either:
  $$\widehat{p_{exact}}(y_{T+1}|\mathbf{y}) = \frac{1}{M}\sum_{i-1}^{M} p(y_{T+1}|\mathbf{y}, \boldsymbol{\theta}^{(i)})$$
  2. or: $\widehat{p_{exact}}(y_{T+1}|\mathbf{y})$ constructed from draws of $y_{T+1}^{(i)}$ simulated from $p(y_{T+1}|\mathbf{y}, \boldsymbol{\theta}^{(i)})$
- i.e. MCMC $\Rightarrow$ **exact Bayesian forecasting**
  - (up to simulation error)

Approximate Bayesian forecasting

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

Approximate Bayesian forecasting

- Can/must resort to **approximate Bayesian inference**
- $\Rightarrow$ **goal** then is to produce **an approximation to** $p(\theta|\mathbf{y})$
- $\Rightarrow$ an **approximation to** $p_{exact}(y_{T+1}|\mathbf{y})$
- **Approximations to** $p(\theta|\mathbf{y})$?
  - 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 $p(\boldsymbol{\theta}|\mathbf{y})$
- 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 $\boldsymbol{\theta}^i$ from $p(\boldsymbol{\theta})$
2. Simulate **pseudo-data** $\mathbf{z}^i$, $i = 1, 2, ..., N$, from $p(\mathbf{z}|\boldsymbol{\theta}^i)$
3. Select $\boldsymbol{\theta}^i$ such that:

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

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

4. Selected draws $\Rightarrow$ simulation-based estimate of $p(\boldsymbol{\theta}|\boldsymbol{\eta}(\mathbf{y}))$

Approximate Bayesian forecasting

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

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

- What is $p_{ABC}(y_{T+1}|\mathbf{y})$ & how does it relate to $p_{exact}(y_{T+1}|\mathbf{y})$?
- We show (in **‘ABF’, 2018**) that:
  - $p_{ABC}(y_{T+1}|\mathbf{y})$ **is** a **proper** density function
  - $p_{ABC}(y_{T+1}|\mathbf{y}) = p_{exact}(y_{T+1}|\mathbf{y})$ iff $\boldsymbol{\eta}(\mathbf{y})$ is **sufficient (!)**
  - $p_{ABC}(y_{T+1}|\mathbf{y}) \approx p_{exact}(y_{T+1}|\mathbf{y})$ even when $\boldsymbol{\eta}(\mathbf{y})$ is **not sufficient**

Furthermore......

- Under **Bayesian consistency** of:
  - $p(\boldsymbol{\theta}|\mathbf{y})$ (standard regularity) and
  - $p(\boldsymbol{\theta}|\eta(\mathbf{y}))$ (**Frazier, Martin, Robert and Rousseau, 2018**)
- the predictive distributions:

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

'**merge**', in the sense that:

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

  - **Blackwell and Dubins (1962)**
- $\Rightarrow$ for large enough T **exact** and (consistent) **ABC-based** predictives are **equivalent!**

Furthermore......

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

$$
\begin{aligned}
E[S(p_{exact}, y_{T+1})] &= \int_{y\in\Omega} S(p_{exact}, y_{T+1}) \underbrace{p(y_{T+1}|\mathbf{y}, \boldsymbol{\theta}_0)}_{p_{truth}} dy_{T+1} \\
&\geq \int_{y\in\Omega} S(p_{ABC}, y_{T+1}) \underbrace{p(y_{T+1}|\mathbf{y}, \boldsymbol{\theta}_0)}_{p_{truth}} dy_{T+1} \\
&= E[S(p_{ABC}, y_{T+1})]
\end{aligned}
$$

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**

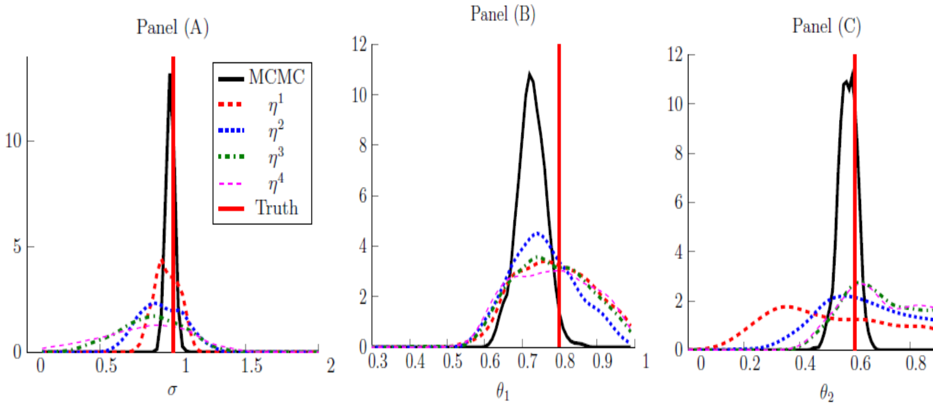
$$\gamma_l = cov(y_t, y_{t-l})$$

- to construct (alternative vectors of) summary statistics:

$$\begin{aligned} \eta^{(1)}(\mathbf{y}) &= (\gamma_0, \gamma_1)'; \ \eta^{(2)}(\mathbf{y}) = (\gamma_0, \gamma_1, \gamma_2)' \\ \eta^{(3)}(\mathbf{y}) &= (\gamma_0, \gamma_1, \gamma_2, \gamma_3)'; \ \eta^{(4)}(\mathbf{y}) = (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4)' \end{aligned}$$

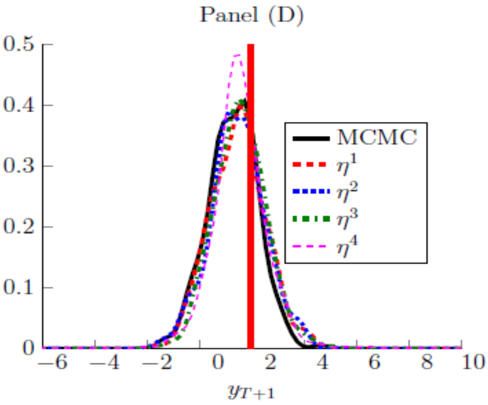
- MA dependence $\Rightarrow$ no reduction to sufficiency possible
  - $\Rightarrow p(\boldsymbol{\theta}|\eta^{(j)}(\mathbf{y})) \neq p(\boldsymbol{\theta}|\mathbf{y})$ for all $j = 1, 2, 3, 4$
  - **What about** $p_{ABC}(y_{T+1}|\mathbf{y})$ *versus* $p_{exact}(y_{T+1}|\mathbf{y})$??

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 $\eta^{(j)}(\mathbf{y})$!!

Expected scores: exact and ABC: T = 500

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

	ABC av. score				Exact av. score
	$\eta^{(1)}(\mathbf{y})$	$\eta^{(2)}(\mathbf{y})$	$\eta^{(3)}(\mathbf{y})$	$\eta^{(4)}(\mathbf{y})$	
LS	-1.43	-1.42	-1.43	-1.43	**-1.40**
QS	0.28	0.28	0.28	0.28	**0.29**
CRPS	-0.57	**-0.56**	-0.57	-0.57	**-0.56**

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

ABC prediction in state space models?

- How does one compute $p_{ABC}(y_{T+1}|\mathbf{y})$ in **state space models**?
  - Does one condition **state inference** only on $\eta(\mathbf{y})$?
- Given a financial return, $y_t = \ln P_t - \ln P_{t-1}$
- Assume **stochastic volatility:**

$$
\begin{aligned}
y_t &= \sqrt{V_t}\varepsilon_t; \qquad \varepsilon_t \sim i.i.d.N(0,1) \\
\ln V_t &= \theta_1 \ln V_{t-1} + \eta_t; \qquad \eta_t \sim i.i.d.N(0,\theta_2)
\end{aligned}
$$

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

ABC prediction in state space models?

- **Exact:**

$$p_{exact}(y_{T+1}|\mathbf{y}) = \int_{V_{T+1}} \int_{\mathbf{V}} \int_{\boldsymbol{\theta}} p(y_{T+1}|V_{T+1}) \times p(V_{T+1}|V_T, \mathbf{y}, \boldsymbol{\theta}) \underbrace{p(\mathbf{V}|\boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}|\mathbf{y})}_{p(\mathbf{V},\boldsymbol{\theta}|\mathbf{y})} d\boldsymbol{\theta} d\mathbf{V} dV_{T+1}$$

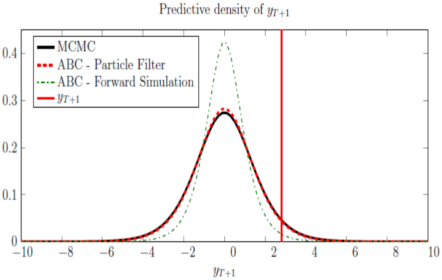
- **MCMC** used to draw from $p(\mathbf{V}, \boldsymbol{\theta}|\mathbf{y})$
- $\Rightarrow$ **independent** draws from $p(V_{T+1}|V_T, \mathbf{y}, \boldsymbol{\theta})$ and $p(y_{T+1}|V_{T+1}) \Rightarrow y_{T+1}^{(i)}$
- $\Rightarrow \widehat{p_{exact}}(y_{T+1}|\mathbf{y})$

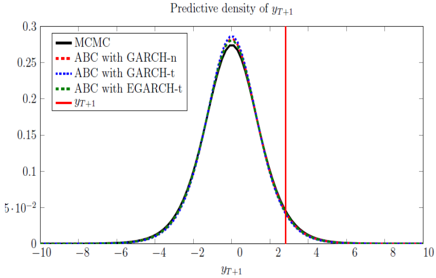
ABC prediction in state space models?

- **ABC:**

$$p_{ABC}(y_{T+1}|\mathbf{y}) = \int_{V_{T+1}} \int_{\mathbf{V}} \int_{\boldsymbol{\theta}} p(y_{T+1}|V_{T+1})$$
$$\times \; p(V_{T+1}|V_T, \mathbf{y}, \boldsymbol{\theta}) p(\mathbf{V}|\boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}|\boldsymbol{\eta}(\mathbf{y})) d\boldsymbol{\theta} d\mathbf{V} dV_{T+1}$$

- **ABC** used to draw from $p(\boldsymbol{\theta}|\boldsymbol{\eta}(\mathbf{y}))$
- (with $\boldsymbol{\eta}(\mathbf{y})$ based on an approximating **auxiliary** GARCH model)
- $\Rightarrow$ **particle filtering** used to integrate out $\mathbf{V}$
- $\Rightarrow$ yields **full posterior inference** (i.e. $|\mathbf{y}$) on V_T
- Exact inference (MCMC) on $\mathbf{V}_{1:T-1}$ not required





- Nature of ABC inference on $\boldsymbol{\theta}$ of little importance.....
- $\Rightarrow$ **all** $p_{ABC}(y_{T+1}|\mathbf{y}) \approx p_{exact}(y_{T+1}|\mathbf{y})$!
- What if condition V_T on $\boldsymbol{\eta}(\mathbf{y})$ 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**: $p(y_{T+1}|V_{T+1})$ and $p(V_{T+1}|V_T, \mathbf{y}, \boldsymbol{\theta})$ right!
- But only need **particle filtering** to do that
- $\Rightarrow$ ABC prediction still based on independent sampling
- $\Rightarrow$ **parallel computing** can still be exploited

Empirical setting??

- Thus far? Have assumed:
  1. That the **DGP:** $p(y_{T+1}, \mathbf{y}|\boldsymbol{\theta}) = p(y_{T+1}|\mathbf{y}, \boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})$ is **correctly specified**
     - (whether latent states are playing a role or not.....)
  2. That we have access to $p(\boldsymbol{\theta}|\mathbf{y}) \Rightarrow p_{exact}(y_{T+1}|\mathbf{y})$
     - for assessment of $p(\boldsymbol{\theta}|\boldsymbol{\eta}(\mathbf{y})) \Rightarrow p_{ABC}(y_{T+1}|\mathbf{y})$
- In a realistic empirical setting:
  1. The assumed **DGP** will be **misspecified**
  2. We are accessing $p_{ABC}(y_{T+1}|\mathbf{y})$ because we cannot (or it is too computationally burdensome) to access $p_{exact}(y_{T+1}|\mathbf{y})$
     - $\Rightarrow$ **no benchmark** for $p_{ABC}(y_{T+1}|\mathbf{y})$
  3. **Critically**.....the sense in which $p_{exact}(y_{T+1}|\mathbf{y})$ remains the **gold standard** is no longer clear

Empirical setting??

- Two routes:

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

2. Choose $\eta(\mathbf{y})$ according to a **predictive criterion**
    - $\Rightarrow \eta(\mathbf{y}) = f^n(S(p, y_{T+1}))$
    - 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, $p_t = \ln P_t$):
- Jump diffusion with (square root) stochastic volatility:

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

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

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

- **Assume:**

$$dp_t = \sqrt{V_t} dB_t^p$$

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

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**
- $p(\boldsymbol{\theta}|\mathbf{y})$ (under regularity) concentrates onto **pseudo-true** $\boldsymbol{\theta}$, $\boldsymbol{\theta}^*$
- where $\boldsymbol{\theta}^*$ is close to $\boldsymbol{\theta}_0$ (in KL-based sense)

Example of route 2

- $\Rightarrow$
$$\lim_{T\rightarrow\infty} p_{exact}(y_{T+1}|\mathbf{y}) = p(y_{T+1}|\mathbf{y},\boldsymbol{\theta}^{*}) = what??$$
- p is misspecified
- $\boldsymbol{\theta}^{*} \neq \boldsymbol{\theta}_0$
- And we have nowhere else to go.....
- With an **ABC-type** approach we have more room to move.....

Apply ABC-type principles: Option 1

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

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

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

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

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

$$\eta^i(.) = \frac{1}{n_e} \sum_{\tau=0}^{n_e} S(p(y_{T+1+\tau}|\mathbf{y}_{1;T+\tau}, \boldsymbol{\theta}^i), y^0_{T+1+\tau})$$

5. Select $\boldsymbol{\theta}^i$ such that:

$$\eta^i(.) \; > \text{the } \textbf{highest } (\alpha\%, \text{say}) \text{ quantile}$$

Apply ABC-type principles: Option 1

- $\Rightarrow$ produces a range of **plausible** $p(y_{T+1+n_e}|\mathbf{y}_{1:T+n_e}, \boldsymbol{\theta}^i)$
- 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:

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

- By computing (over N_a 'accepted' $\boldsymbol{\theta}^i$):

$$p_{av}(y_{T+1+n_e}|\mathbf{y}_{1:T+n_e}) = \frac{1}{N_a}\sum_{i=1}^{N_a} p(y_{T+1+n_e}|\mathbf{y}_{1:T+n_e}, \boldsymbol{\theta}^i)$$

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** $q(y_{T+1}, \mathbf{y}, \boldsymbol{\beta})$ that *approximates* $p(y_{T+1}, \mathbf{y}, \boldsymbol{\theta})$
- $\widehat{\boldsymbol{\beta}}_{MLE} \Rightarrow \boldsymbol{\eta}(\mathbf{y})$
- Aim in auxiliary-model based ABC for **inference**?
- Choose $q(y_{T+1}, \mathbf{y}, \boldsymbol{\theta})$ to capture features of $p(y_{T+1}, \mathbf{y}, \boldsymbol{\theta})$
- If $q(y_{T+1}, \mathbf{y}, \boldsymbol{\theta})$ 'nests' (a **correctly specified**) $p(y_{T+1}, \mathbf{y}, \boldsymbol{\theta})$
  - $\Rightarrow \boldsymbol{\eta}(\mathbf{y}) = \widehat{\boldsymbol{\beta}}_{MLE}$ is **asymptotically sufficient** for $\boldsymbol{\theta}$
  - $\Rightarrow p(\boldsymbol{\theta}|\boldsymbol{\eta}(\mathbf{y})) = p(\boldsymbol{\theta}|\mathbf{y})$ (for large T)
  - $\Rightarrow$ 'ideal' $q(y_{T+1}, \mathbf{y}, \boldsymbol{\theta})$ is **highly parameterized**

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....
- $\Rightarrow$ Approach in auxiliary-model based ABC for **forecasting**?
- Pick a **simple parsimoneous 'auxiliary predictive':**

$$q(y_{T+1}|\mathbf{y}_{1:T}, \boldsymbol{\beta})$$

- And **select** $\boldsymbol{\theta}^i$ (and, hence, $p(y_{T+1}|\mathbf{y}, \boldsymbol{\theta}^i)$)
- such that the predictive performance of $p(y_{T+1}|\mathbf{y}, \boldsymbol{\theta}^i)$ **matches** that of $q(y_{T+1}|\mathbf{y}_{1:T}, \boldsymbol{\beta})$
- $\Rightarrow$ Replace Steps 4. and 5. above with:

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

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

5. Select $\boldsymbol{\theta}^i$ such that:

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

- Produces a simulation-based estimate of a **different**:

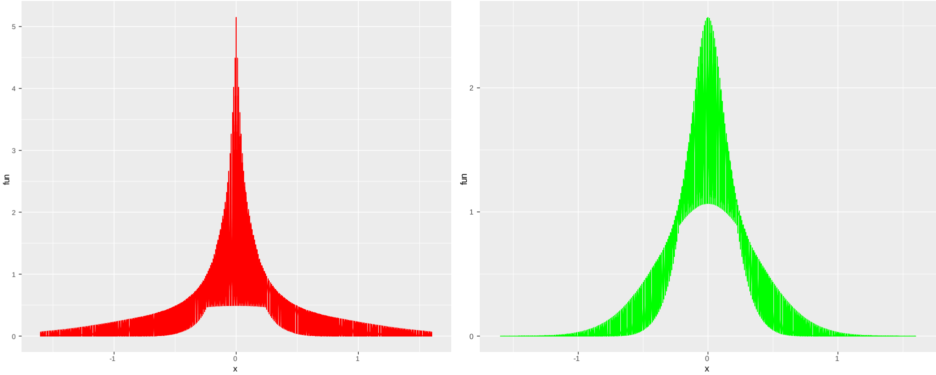
$$p_{ABC}(y_{T+1+n_e} | \mathbf{y}_{1:T+n_e}) = \int p(y_{T+1+n_e} | \mathbf{y}_{1:T+n_e}, \boldsymbol{\theta}) p(\boldsymbol{\theta} | \eta(.)) d\boldsymbol{\theta}$$

- in which $\eta(.)$ reflects **a different** measure of predictive perfomance

(Very!) preliminary results

- Choose $q(y_{T+1}|\mathbf{y}_{1:T}, \boldsymbol{\beta})$ 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 $p_{ABC}(y_{T+1+n_e}|\mathbf{y}_{1:T+n_e})$
- Roll the whole process forward:
  - Compute **log scores** for 25 one-step-ahead predictions for both estimates of $p_{ABC}(y_{T+1+n_e}|\mathbf{y}_{1:T+n_e})$

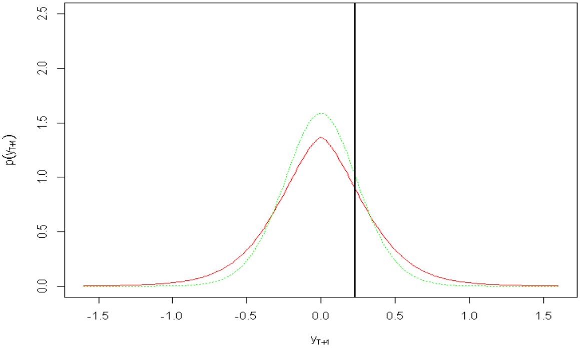
Plots of accepted conditional predictives



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

Averaged accepted predictives

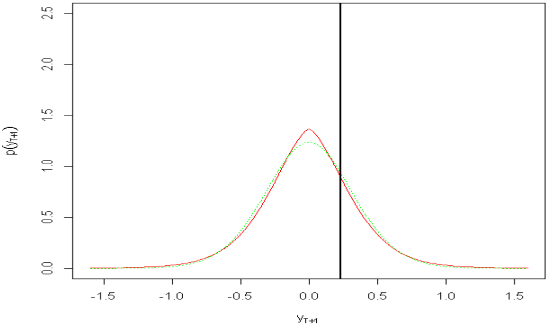
- Or **average**: to produce estimates of $p_{ABC}(y_{T+1+n_e}|\mathbf{y}_{1:T+n_e})$:



- Median scores (over 25 one-step-ahead periods):
  - **Option 1**: -0.262; **Option 2**: -0.114

Change the auxiliary predictive?

- Choose $q(y_{T+1+\tau}|\mathbf{y}_{1:T+\tau}, \boldsymbol{\beta})$ 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

To come.....

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

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

- where $\boldsymbol{\theta}^*$ minimizes the KL divergence of the assumed model from the **true DGP**
- Will $p_{exact}(y_{T+1}|\mathbf{y})$ still 'win' in terms of **log score**?
- But $p_{ABC}(y_{T+1}|\mathbf{y})$ 'win' in terms of **alternative performance criteria** (that have informed $\eta(.)$)?

To come....

- If so
- $\Rightarrow$ Ideas may have relevance **beyond** usual ABC scenario
- $\Rightarrow$ 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 ε in:
  - **Frazier, Martin, Robert and Rousseau, 2018**
- $\cdots\cdots$ all in good time.....