Stochastic Model Specification Search for Gaussian and Non-Gaussian State Space Models

by Sylvia Frühwirth-Schnatter and Helga Wagner

Discussion by Gabriele Fiorentini

University of Florence
and
Rimini Centre for Economic Analysis (RCEA)

- The paper is enjoyable and instructive
- In a nutshell: "the authors propose an ingenious parameterization of STS models that fits the variable selection approach of George and McCulloch"

- The paper is enjoyable and instructive
- In a nutshell: "the authors propose an ingenious parameterization of STS models that fits the variable selection approach of George and McCulloch"

Outline of Discussion

- My view on the importance of Bayesian model discrimination in STS models
- The noncentered prameterization and a "classical" perspective
- Implementation of George and McCulloch variable selection method
- Motivations for stochastic model specification search in the STS framework

Outline of Discussion

- My view on the importance of Bayesian model discrimination in STS models
- The noncentered prameterization and a "classical" perspective
- Implementation of George and McCulloch variable selection method
- Motivations for stochastic model specification search in the STS framework

Outline of Discussion

- My view on the importance of Bayesian model discrimination in STS models
- The noncentered prameterization and a "classical" perspective
- Implementation of George and McCulloch variable selection method
- Motivations for stochastic model specification search in the STS framework

Outline of Discussion

- My view on the importance of Bayesian model discrimination in STS models
- The noncentered prameterization and a "classical" perspective
- Implementation of George and McCulloch variable selection method
- Motivations for stochastic model specification search in the STS framework

Bayesian model discrimination in STS models

- **STS models** widely used in applied macroeconomics.
- Ideally macroeconomics should guide the time series model choice. Economic theories however rarely point to a unique specification, and discriminating between different possibilities can be difficult.
- Even in a simple trend plus noise decomposition the amount of uncertainty may be huge
  a) What model for the trend?
  b) Deviations from trend should be a pure noise or a stationary autocorrelated short term component ?
  c) Identification of the component models

1) Model selection 2) Model identification

Bayesian model discrimination in STS models

- **STS models** widely used in applied macroeconomics.
- Ideally macroeconomics should guide the time series model choice. Economic theories however rarely point to a unique specification, and discriminating between different possibilities can be difficult.
- Even in a simple trend plus noise decomposition the amount of uncertainty may be huge
  a) What model for the trend?
  b) Deviations from trend should be a pure noise or a stationary autocorrelated short term component ?
  c) Identification of the component models

1) Model selection   2) Model identification

Bayesian model discrimination in STS models

- **STS models** widely used in applied macroeconomics.
- Ideally macroeconomics should guide the time series model choice. Economic theories however rarely point to a unique specification, and discriminating between different possibilities can be difficult.
- Even in a simple trend plus noise decomposition the amount of uncertainty may be huge
  a) What model for the trend?
  b) Deviations from trend should be a pure noise or a stationary autocorrelated short term component ?
  c) Identification of the component models

1) Model selection 2) Model identification

Bayesian model discrimination in STS models

- **STS models** widely used in applied macroeconomics.
- Ideally macroeconomics should guide the time series model choice. Economic theories however rarely point to a unique specification, and discriminating between different possibilities can be difficult.
- Even in a simple trend plus noise decomposition the amount of uncertainty may be huge
  a) What model for the trend?
  b) Deviations from trend should be a pure noise or a stationary autocorrelated short term component ?
  c) Identification of the component models

1) Model selection **2)** Model identification

Model selection

The Bayesian's answer to **model selection** is the posterior probability of model M_i:

$$p(M_i|y) \propto f(y|M_i)p(M_i)$$

Let $j = \arg\max_i \; p(M_i|y)$, then you can take $f(trend|y, M_j)$

I like more $f(trend|y)$ which reflects the uncertainty faced by the researcher.

If $p(M_j|y)$ is very large then $f(trend|y) \approx f(trend|y, M_j)$

Model selection

The Bayesian's answer to **model selection** is the posterior probability of model M_i:

$$p(M_i|y) \propto f(y|M_i)p(M_i)$$

Let $j = \arg\max_i \; p(M_i|y)$, then you can take $f(trend|y, M_j)$

I like more $f(trend|y)$ which reflects the uncertainty faced by the researcher.

If $p(M_j|y)$ is very large then $f(trend|y) \approx f(trend|y, M_j)$

Model selection

The Bayesian's answer to **model selection** is the posterior probability of model M_i:

$$p(M_i|y) \propto f(y|M_i)p(M_i)$$

Let $j = \arg\max_i \; p(M_i|y)$, then you can take $f(trend|y, M_j)$

I like more $f(trend|y)$ which reflects the uncertainty faced by the researcher.

If $p(M_j|y)$ is very large then $f(trend|y) \approx f(trend|y, M_j)$

Model selection

The Bayesian's answer to **model selection** is the posterior probability of model M_i:

$$p(M_i|y) \propto f(y|M_i)p(M_i)$$

Let $j = \arg\max_i \; p(M_i|y)$, then you can take $f(trend|y, M_j)$

I like more $f(trend|y)$ which reflects the uncertainty faced by the researcher.

If $p(M_j|y)$ is very large then $f(trend|y) \approx f(trend|y, M_j)$

Model selection

The Bayesian's answer to **model selection** is the posterior probability of model M_i:

$$p(M_i|y) \propto f(y|M_i)p(M_i)$$

Let $j = \arg\max_i \; p(M_i|y)$, then you can take $f(trend|y, M_j)$

I like more $f(trend|y)$ which reflects the uncertainty faced by the researcher.

If $p(M_j|y)$ is very large then $f(trend|y) \approx f(trend|y, M_j)$

Model Identification

Let M_i be a simple trend plus noise model in which the trend is an I(1) process without drift

$$y_t = \mu_t + \epsilon_t$$

$$\mu_t = \mu_{t-1} + \omega_t + \vartheta\omega_{t-1} \qquad 0 \le \vartheta \le 1$$

there are infinitely many combinations of ϑ and the signal to noise ratio that are o.e.

Instead of imposing a specific identifying restriction on ϑ we could specify a prior distribution $f(\vartheta)$.

Then $f(trend|y, M_i)$ would reflect our uncertainty.

Notice that in the above discussion I assume $\epsilon_t \perp \omega_t$.

Model Identification

Let M_i be a simple trend plus noise model in which the trend is an I(1) process without drift

$$y_t = \mu_t + \epsilon_t$$

$$\mu_t = \mu_{t-1} + \omega_t + \vartheta\omega_{t-1} \qquad 0 \leq \vartheta \leq 1$$

there are infinitely many combinations of ϑ and the signal to noise ratio that are o.e.

Instead of imposing a specific identifying restriction on ϑ we could specify a prior distribution $f(\vartheta)$.

Then $f(trend|y, M_i)$ would reflect our uncertainty.

Notice that in the above discussion I assume $\epsilon_t \perp \omega_t$.

Model Identification

Let M_i be a simple trend plus noise model in which the trend is an I(1) process without drift

$$y_t = \mu_t + \epsilon_t$$

$$\mu_t = \mu_{t-1} + \omega_t + \vartheta\omega_{t-1} \qquad 0 \leq \vartheta \leq 1$$

there are infinitely many combinations of ϑ and the signal to noise ratio that are o.e.

Instead of imposing a specific identifying restriction on ϑ we could specify a prior distribution $f(\vartheta)$.

Then $f(trend|y, M_i)$ would reflect our uncertainty.

Notice that in the above discussion I assume $\epsilon_t \perp \omega_t$.

Model Identification

Let M_i be a simple trend plus noise model in which the trend is an I(1) process without drift

$$y_t = \mu_t + \epsilon_t$$

$$\mu_t = \mu_{t-1} + \omega_t + \vartheta\omega_{t-1} \qquad 0 \leq \vartheta \leq 1$$

there are infinitely many combinations of ϑ and the signal to noise ratio that are o.e.

Instead of imposing a specific identifying restriction on ϑ we could specify a prior distribution $f(\vartheta)$.

Then $f(trend|y, M_i)$ would reflect our uncertainty.

Notice that in the above discussion I assume $\epsilon_t \perp \omega_t$.

Model Identification

Let M_i be a simple trend plus noise model in which the trend is an I(1) process without drift

$$y_t = \mu_t + \epsilon_t$$

$$\mu_t = \mu_{t-1} + \omega_t + \vartheta\omega_{t-1} \qquad 0 \leq \vartheta \leq 1$$

there are infinitely many combinations of ϑ and the signal to noise ratio that are o.e.

Instead of imposing a specific identifying restriction on ϑ we could specify a prior distribution $f(\vartheta)$.

Then $f(trend|y, M_i)$ would reflect our uncertainty.

Notice that in the above discussion I assume $\epsilon_t \perp \omega_t$.

Model Identification

Let M_i be a simple trend plus noise model in which the trend is an I(1) process without drift

$$y_t = \mu_t + \epsilon_t$$

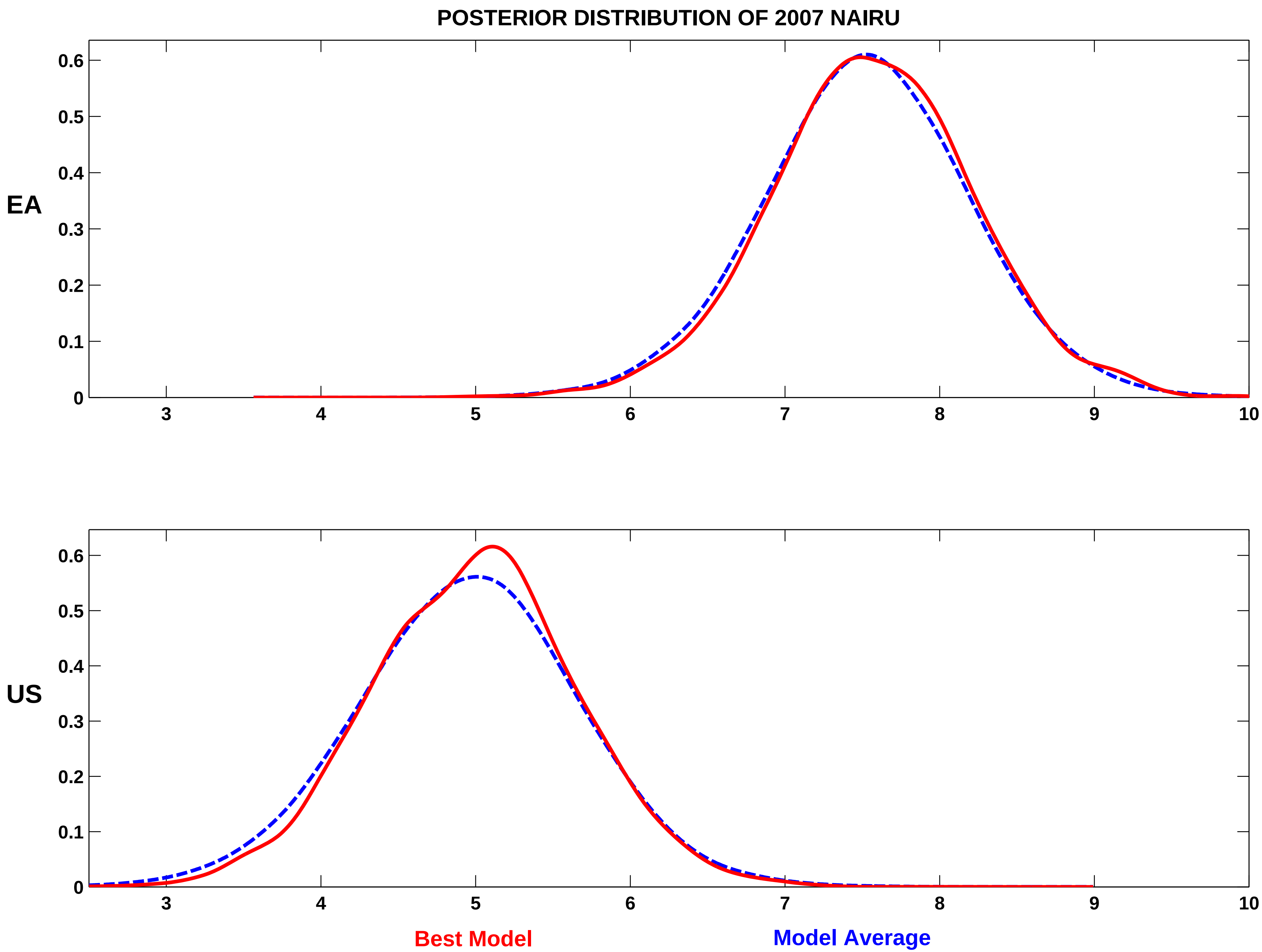
$$\mu_t = \mu_{t-1} + \omega_t + \vartheta\omega_{t-1} \qquad 0 \leq \vartheta \leq 1$$

there are infinitely many combinations of ϑ and the signal to noise ratio that are o.e.

Instead of imposing a specific identifying restriction on ϑ we could specify a prior distribution $f(\vartheta)$.

Then $f(trend|y, M_i)$ would reflect our uncertainty.

Notice that in the above discussion I assume $\epsilon_t \perp \omega_t$.


POSTERIOR DISTRIBUTION OF 2007 NAIRU
EA
US
0.6
0.5
0.4
0.3
0.2
0.1
0
3
4
5
6
7
8
9
10
Best Model
Model Average

The Noncentered parameterization

The parameterization $\pm\sqrt{\theta}$ is discussed at length.
It is argued that:

- Normal prior more suitable than IG under model uncertainty. Results are less sensitive to prior specification as the normal prior is less influential than IG prior.

  For θ is re-scaled χ^2 versus IG...

- Better mixing properties

  If mixing comes from sign switch...

- Bring important information about the hypothesis $\theta = 0$

  It should help also under a classical approach...

It seems odd that the parameterization is not used for σ_ϵ^2
Also, the normal prior on $\sqrt{\theta}$ is actually NIG on $(\sqrt{\theta}, \sigma_\epsilon^2)$

The Noncentered parameterization

The parameterization $\pm\sqrt{\theta}$ is discussed at length.
It is argued that:

- Normal prior more suitable than IG under model uncertainty. Results are less sensitive to prior specification as the normal prior is less influential than IG prior.

  For θ is re-scaled χ^2 versus IG...

- Better mixing properties

  If mixing comes from sign switch...

- Bring important information about the hypothesis $\theta = 0$

  It should help also under a classical approach...

It seems odd that the parameterization is not used for σ_ϵ^2
Also, the normal prior on $\sqrt{\theta}$ is actually NIG on $(\sqrt{\theta}, \sigma_\epsilon^2)$

The Noncentered parameterization

The parameterization $\pm\sqrt{\theta}$ is discussed at length.
It is argued that:

- Normal prior more suitable than IG under model uncertainty. Results are less sensitive to prior specification as the normal prior is less influential than IG prior.

  For θ is re-scaled χ^2 versus IG...

- Better mixing properties

  If mixing comes from sign switch...

- Bring important information about the hypothesis $\theta = 0$

  It should help also under a classical approach...

It seems odd that the parameterization is not used for σ_ϵ^2
Also, the normal prior on $\sqrt{\theta}$ is actually NIG on $(\sqrt{\theta}, \sigma_\epsilon^2)$

The Noncentered parameterization

The parameterization $\pm\sqrt{\theta}$ is discussed at length.
It is argued that:

- Normal prior more suitable than IG under model uncertainty. Results are less sensitive to prior specification as the normal prior is less influential than IG prior.

  For θ is re-scaled χ^2 versus IG...

- Better mixing properties

  If mixing comes from sign switch...

- Bring important information about the hypothesis $\theta = 0$

  It should help also under a classical approach...

It seems odd that the parameterization is not used for σ_ϵ^2
Also, the normal prior on $\sqrt{\theta}$ is actually NIG on $(\sqrt{\theta}, \sigma_\epsilon^2)$

The Noncentered parameterization

The parameterization $\pm\sqrt{\theta}$ is discussed at length.
It is argued that:

- Normal prior more suitable than IG under model uncertainty. Results are less sensitive to prior specification as the normal prior is less influential than IG prior.
  For θ is re-scaled χ^2 versus IG...

- Better mixing properties
  If mixing comes from sign switch...

- Bring important information about the hypothesis $\theta = 0$
  It should help also under a classical approach...

It seems odd that the parameterization is not used for σ_ϵ^2
Also, the normal prior on $\sqrt{\theta}$ is actually NIG on $(\sqrt{\theta}, \sigma_\epsilon^2)$

The Noncentered parameterization

The parameterization $\pm\sqrt{\theta}$ is discussed at length.
It is argued that:

- Normal prior more suitable than IG under model uncertainty. Results are less sensitive to prior specification as the normal prior is less influential than IG prior.

  For θ is re-scaled χ^2 versus IG...

- Better mixing properties

  If mixing comes from sign switch...

- Bring important information about the hypothesis $\theta = 0$

  It should help also under a classical approach...

It seems odd that the parameterization is not used for σ^2_ϵ
Also, the normal prior on $\sqrt{\theta}$ is actually NIG on $(\sqrt{\theta}, \sigma^2_\epsilon)$

The Noncentered parameterization

The parameterization $\pm\sqrt{\theta}$ is discussed at length.
It is argued that:

- Normal prior more suitable than IG under model uncertainty. Results are less sensitive to prior specification as the normal prior is less influential than IG prior.

  For θ is re-scaled χ^2 versus IG...

- Better mixing properties

  If mixing comes from sign switch...

- Bring important information about the hypothesis $\theta = 0$

  It should help also under a classical approach...

It seems odd that the parameterization is not used for σ_ϵ^2
Also, the normal prior on $\sqrt{\theta}$ is actually NIG on $(\sqrt{\theta}, \sigma_\epsilon^2)$

The Noncentered parameterization

The parameterization $\pm\sqrt{\theta}$ is discussed at length.
It is argued that:

- Normal prior more suitable than IG under model uncertainty. Results are less sensitive to prior specification as the normal prior is less influential than IG prior.

  For θ is re-scaled χ^2 versus IG...

- Better mixing properties

  If mixing comes from sign switch...

- Bring important information about the hypothesis $\theta = 0$

  It should help also under a classical approach...

It seems odd that the parameterization is not used for σ_ϵ^2
Also, the normal prior on $\sqrt{\theta}$ is actually NIG on $(\sqrt{\theta}, \sigma_\epsilon^2)$

The Noncentered parameterization

The parameterization $\pm\sqrt{\theta}$ is discussed at length.
It is argued that:

- Normal prior more suitable than IG under model uncertainty. Results are less sensitive to prior specification as the normal prior is less influential than IG prior.

  For θ is re-scaled χ^2 versus IG...

- Better mixing properties

  If mixing comes from sign switch...

- Bring important information about the hypothesis $\theta = 0$

  It should help also under a classical approach...

It seems odd that the parameterization is not used for σ_ϵ^2
Also, the normal prior on $\sqrt{\theta}$ is actually NIG on $(\sqrt{\theta}, \sigma_\epsilon^2)$

The Noncentered parameterization

The parameterization $\pm\sqrt{\theta}$ is discussed at length.
It is argued that:

- Normal prior more suitable than IG under model uncertainty. Results are less sensitive to prior specification as the normal prior is less influential than IG prior.

  For θ is re-scaled χ^2 versus IG...

- Better mixing properties

  If mixing comes from sign switch...

- Bring important information about the hypothesis $\theta = 0$

  It should help also under a classical approach...

It seems odd that the parameterization is not used for σ_ϵ^2
Also, the normal prior on $\sqrt{\theta}$ is actually NIG on $(\sqrt{\theta}, \sigma_\epsilon^2)$

The Noncentered parameterization

The parameterization $\pm\sqrt{\theta}$ is discussed at length.
It is argued that:

- Normal prior more suitable than IG under model uncertainty. Results are less sensitive to prior specification as the normal prior is less influential than IG prior.

  For θ is re-scaled χ^2 versus IG...

- Better mixing properties

  If mixing comes from sign switch...

- Bring important information about the hypothesis $\theta = 0$

  It should help also under a classical approach...

It seems odd that the parameterization is not used for σ_ϵ^2
Also, the normal prior on $\sqrt{\theta}$ is actually NIG on $(\sqrt{\theta}, \sigma_\epsilon^2)$

The Noncentered parameterization and classical statistics

What if one uses the same reparameterization under a standard ML approach.

This would change the constrained maximization of $L(\theta, \ldots)$ into the unconstrained maximization of $L(\phi, \ldots)$ (with $\phi^2 = \theta$).

Maybe, it would also transform a non standard testing problem into a standard one.

It doesn't work as $\frac{\partial L(\phi,\ldots)}{\partial \phi} \mid_{\phi=0} =$ 0 by construction.

Still the likelihood is bimodal when the variance is different from zero and unimodal otherwise

The Noncentered parameterization and classical statistics

What if one uses the same reparameterization under a standard ML approach.

This would change the constrained maximization of $L(\theta, \ldots)$ into the unconstrained maximization of $L(\phi, \ldots)$ (with $\phi^2 = \theta$).

Maybe, it would also transform a non standard testing problem into a standard one.

It doesn't work as $\frac{\partial L(\phi,\ldots)}{\partial \phi}\,|_{\phi=0} = 0$ by construction.

Still the likelihood is bimodal when the variance is different from zero and unimodal otherwise

The Noncentered parameterization and classical statistics

What if one uses the same reparameterization under a standard ML approach.

This would change the constrained maximization of $L(\theta, \ldots)$ into the unconstrained maximization of $L(\phi, \ldots)$ (with $\phi^2 = \theta$).

Maybe, it would also transform a non standard testing problem into a standard one.

It doesn't work as $\frac{\partial L(\phi,\ldots)}{\partial \phi}\big|_{\phi=0} = 0$ by construction.

Still the likelihood is bimodal when the variance is different from zero and unimodal otherwise

The Noncentered parameterization and classical statistics

What if one uses the same reparameterization under a standard ML approach.

This would change the constrained maximization of $L(\theta, \ldots)$ into the unconstrained maximization of $L(\phi, \ldots)$ (with $\phi^2 = \theta$).

Maybe, it would also transform a non standard testing problem into a standard one.

It doesn't work as $\frac{\partial L(\phi,\ldots)}{\partial \phi}\,|_{\phi=0} = 0$ by construction.

Still the likelihood is bimodal when the variance is different from zero and unimodal otherwise

The Noncentered parameterization and classical statistics

What if one uses the same reparameterization under a standard ML approach.

This would change the constrained maximization of $L(\theta, \ldots)$ into the unconstrained maximization of $L(\phi, \ldots)$ (with $\phi^2 = \theta$).

Maybe, it would also transform a non standard testing problem into a standard one.

It doesn't work as $\frac{\partial L(\phi, \ldots)}{\partial \phi} |_{\phi=0} =$ 0 by construction.

Still the likelihood is bimodal when the variance is different from zero and unimodal otherwise

The Noncentered parameterization and classical statistics

What if one uses the same reparameterization under a standard ML approach.

This would change the constrained maximization of $L(\theta, \ldots)$ into the unconstrained maximization of $L(\phi, \ldots)$ (with $\phi^2 = \theta$).

Maybe, it would also transform a non standard testing problem into a standard one.

It doesn't work as $\frac{\partial L(\phi,\ldots)}{\partial \phi} \mid_{\phi=0} = 0$ by construction.

Still the likelihood is bimodal when the variance is different from zero and unimodal otherwise

Implementation of the variable selection approach

There are different ways to implement George & McCulloch

The one adopted here requires the use of conjugate priors and when an indicator is equal to zero sets to zero also the corresponding parameter.

Could we use nonconjugate priors?

Does it make sense to set a normal prior with very small variance on $\sqrt{\theta}$ when the corresponding indicator is zero? Otherwise care must be exercised not the get reducible chains.

The method would clearly work for a trend+cycle+noise decomposition

But would it work for an exact trend+cycle decomposition?

Implementation of the variable selection approach

There are different ways to implement George & McCulloch

The one adopted here requires the use of conjugate priors and when an indicator is equal to zero sets to zero also the corresponding parameter.

Could we use nonconjugate priors?

Does it make sense to set a normal prior with very small variance on $\sqrt{\theta}$ when the corresponding indicator is zero? Otherwise care must be exercised not the get reducible chains.

The method would clearly work for a trend+cycle+noise decomposition

But would it work for an exact trend+cycle decomposition?

Implementation of the variable selection approach

There are different ways to implement George & McCulloch

The one adopted here requires the use of conjugate priors and when an indicator is equal to zero sets to zero also the corresponding parameter.

Could we use nonconjugate priors?

Does it make sense to set a normal prior with very small variance on $\sqrt{\theta}$ when the corresponding indicator is zero? Otherwise care must be exercised not the get reducible chains.

The method would clearly work for a trend+cycle+noise decomposition

But would it work for an exact trend+cycle decomposition?

Implementation of the variable selection approach

There are different ways to implement George & McCulloch

The one adopted here requires the use of conjugate priors and when an indicator is equal to zero sets to zero also the corresponding parameter.

Could we use nonconjugate priors?

Does it make sense to set a normal prior with very small variance on $\sqrt{\theta}$ when the corresponding indicator is zero? Otherwise care must be exercised not the get reducible chains.

The method would clearly work for a trend+cycle+noise decomposition

But would it work for an exact trend+cycle decomposition?

Implementation of the variable selection approach

There are different ways to implement George & McCulloch

The one adopted here requires the use of conjugate priors and when an indicator is equal to zero sets to zero also the corresponding parameter.

Could we use nonconjugate priors?

Does it make sense to set a normal prior with very small variance on $\sqrt{\theta}$ when the corresponding indicator is zero? Otherwise care must be exercised not the get reducible chains.

The method would clearly work for a trend+cycle+noise decomposition

But would it work for an exact trend+cycle decomposition?

Implementation of the variable selection approach

There are different ways to implement George & McCulloch

The one adopted here requires the use of conjugate priors and when an indicator is equal to zero sets to zero also the corresponding parameter.

Could we use nonconjugate priors?

Does it make sense to set a normal prior with very small variance on $\sqrt{\theta}$ when the corresponding indicator is zero? Otherwise care must be exercised not the get reducible chains.

The method would clearly work for a trend+cycle+noise decomposition

But would it work for an exact trend+cycle decomposition?

Implementation of the variable selection approach

There are different ways to implement George & McCulloch

The one adopted here requires the use of conjugate priors and when an indicator is equal to zero sets to zero also the corresponding parameter.

Could we use nonconjugate priors?

Does it make sense to set a normal prior with very small variance on $\sqrt{\theta}$ when the corresponding indicator is zero? Otherwise care must be exercised not the get reducible chains.

The method would clearly work for a trend+cycle+noise decomposition

But would it work for an exact trend+cycle decomposition?

Stochastic model search vs. marginal likelihood computation

When the model space is large or moderate stochastic model search is the only solution.

There is no need to compute the whole posterior distribution

The stochastic search can work reasonably well even with chains of length smaller than the number of alternative models.

When the model space is small:

$$p(M_i|y) = \frac{f(y|M_i)p(M_i)}{\sum_{j=1}^{K} f(y|M_j)p(M_j)}$$

For instance in the Basic Structural Model I would compute all the marginal likelihoods with the Dickey method virtually without approximation error (Fiorentini, Planas and Rossi, 2008).

Stochastic model search vs. marginal likelihood computation

When the model space is large or moderate stochastic model search is the only solution.

There is no need to compute the whole posterior distribution

The stochastic search can work reasonably well even with chains of length smaller than the number of alternative models.

When the model space is small:

$$p(M_i|y) = \frac{f(y|M_i)p(M_i)}{\sum_{j=1}^{K} f(y|M_j)p(M_j)}$$

For instance in the Basic Structural Model I would compute all the marginal likelihoods with the Dickey method virtually without approximation error (Fiorentini, Planas and Rossi, 2008).

Stochastic model search vs. marginal likelihood computation

When the model space is large or moderate stochastic model search is the only solution.

There is no need to compute the whole posterior distribution

The stochastic search can work reasonably well even with chains of length smaller than the number of alternative models.

When the model space is small:

$$p(M_i|y) = \frac{f(y|M_i)p(M_i)}{\sum_{j=1}^{K} f(y|M_j)p(M_j)}$$

For instance in the Basic Structural Model I would compute all the marginal likelihoods with the Dickey method virtually without approximation error (Fiorentini, Planas and Rossi, 2008).

Stochastic model search vs. marginal likelihood computation

When the model space is large or moderate stochastic model search is the only solution.

There is no need to compute the whole posterior distribution

The stochastic search can work reasonably well even with chains of length smaller than the number of alternative models.

When the model space is small:

$$p(M_i|y) = \frac{f(y|M_i)p(M_i)}{\sum_{j=1}^{K} f(y|M_j)p(M_j)}$$

For instance in the Basic Structural Model I would compute all the marginal likelihoods with the Dickey method virtually without approximation error (Fiorentini, Planas and Rossi, 2008).

Stochastic model search vs. marginal likelihood computation

When the model space is large or moderate stochastic model search is the only solution.

There is no need to compute the whole posterior distribution

The stochastic search can work reasonably well even with chains of length smaller than the number of alternative models.

When the model space is small:

$$p(M_i|y) = \frac{f(y|M_i)p(M_i)}{\sum_{j=1}^{K} f(y|M_j)p(M_j)}$$

For instance in the Basic Structural Model I would compute all the marginal likelihoods with the Dickey method virtually without approximation error (Fiorentini, Planas and Rossi, 2008).

Stochastic model search vs. marginal likelihood computation

When the model space is large or moderate stochastic model search is the only solution.

There is no need to compute the whole posterior distribution

The stochastic search can work reasonably well even with chains of length smaller than the number of alternative models.

When the model space is small:

$$p(M_i|y) = \frac{f(y|M_i)p(M_i)}{\sum_{j=1}^{K} f(y|M_j)p(M_j)}$$

For instance in the Basic Structural Model I would compute all the marginal likelihoods with the Dickey method virtually without approximation error (Fiorentini, Planas and Rossi, 2008).

Stochastic model search: extension

What if we have an unknown number of structural breaks at unknown dates.

The classic marginal likelihood approach is infeasible.

Adapting the stochastic model search to this case would be great.