Essays in statistics and econometrics

Kjartan Kloster Osmundsen

University of Stavanger

PhD defence

Stavanger, June 10th, 2020

Paper list

- **Paper I:** *Using expected shortfall for credit risk regulation* (2018)
  - Published
    Journal of International Financial Markets, Institutions and Money
- **Paper II:** *MCMC for Markov-switching models - Gibbs sampling vs. marginalized likelihood* (2019)
  - Published
    Communications in Statistics - Simulation and Computation
- **Paper III:** *Importance Sampling-based Transport Map Hamiltonian Monte Carlo for Bayesian Hierarchical Models* (2019)
  - Submitted for publication
    Journal of Computational and Graphical Statistics
- **Paper IV:** *Estimating the Competitive Storage Model with Stochastic Trends in Commodity Prices* (2020)
  - Submitted for publication
    Journal of Applied Econometrics

Overview

Main focus

	Paper I	Paper II	Paper III	Paper IV
Application				
Method				

Common themes

	Paper I	Paper II	Paper III	Paper IV
Economic application				
Bayesian inference				
Latent variables				
Hamiltonian Monte Carlo				

Paper I

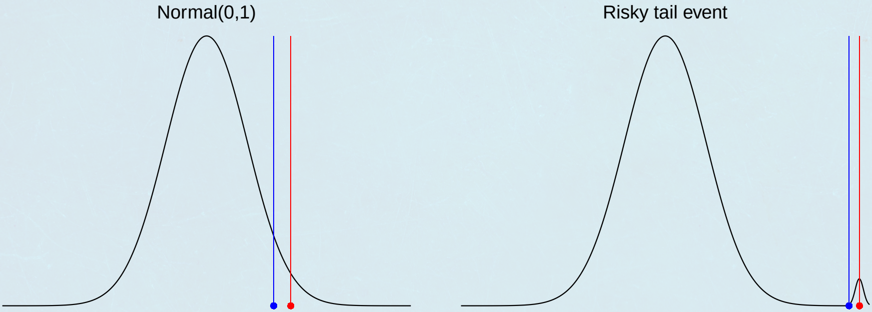
- Osmundsen, Kjartan Kloster (2018). **Using expected shortfall for credit risk regulation**.

Motivation

- In 2016, the *Basel Committee on Banking Supervision* published revised regulatory standards for **market risk**, which include a shift from value at risk (VaR) to expected shortfall (ES) as the underlying risk measure.
  - "A number of weaknesses have been identified with VaR, including its inability to capture tail risk."
- The Committee has so far not considered doing the same for **credit risk** regulation.
  - In fact, ES is turned down for other regulation frameworks to keep consistency with the credit risk framework.
- The Committee argues that ES might be too unstable at the high confidence levels used for credit risk.

Value at risk vs. expected shortfall

- For a given confidence level α, VaR is simply the α-quantile of the loss distribution.
- For continuous distributions, ES is the mean of the loss values greater than or equal to the VaR: $\mathrm{ES}_\alpha(L) = E\left[L | L \geq \mathrm{VaR}_\alpha(L)\right]$.



95 % VaR and 95 % ES.

Value at risk vs. expected shortfall

- VaR does not always reflect the positive effect of diversification.
- Consider a loan of $1 million, with 1 % default probability. Assume the bank's potential profit is $20000, while the whole amount is lost in case of default.

- Alternative 1:
  - The bank offers such loans to 100 independent firms.
  - The bank loses money if two or more firms default (26 %).

	95 % VaR	95 % ES
100 loans	$1.06 million	$1.52 million

- Alternative 2:
  - The bank offers a single large loan of $100 million to a single firm.

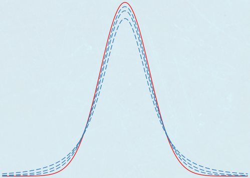
	95 % VaR	95 % ES
1 loan	-$2.0 million	$18.4 million

Confidence level

- Comparisons of VaR and ES is often conducted at the same confidence level.
  - This is not sensible in a capital regulation context.
  - Can lead to wrong conclusions.
- The Basel Committee's capital requirement function is based on a 99.9 % VaR.
  - Partly to protect against inevitable estimation error in the banks' internal models.
- A 99.75 % ES version of the capital requirement function results in a similar capital level.

Parameter sensitivity and estimation uncertainty

- A simulation experiment was conducted to compare the parameter sensitivity of the VaR function and the ES function.
  - Similar for both functions. Small *PD* values give the largest deviations (up to 3%), all favourable for ES



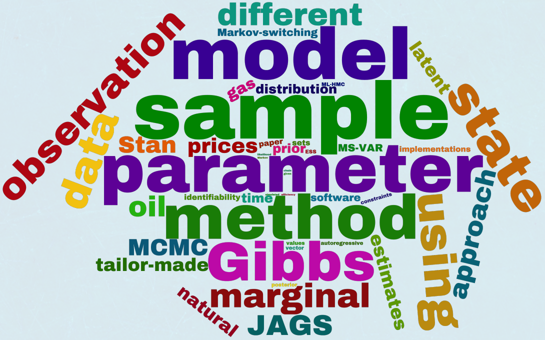
- Using real risk parameter data from a Norwegian savings bank's corporate portfolio, the estimation uncertainty of VaR and ES is simulated for loss distributions with different tail weights.
  - Similar for VaR and ES.

Paper I - Conclusion

- ES has a better ability to accurately capture tail risk, in addition to always reflecting the positive effect of diversification.
  - These shortcomings of VaR have limited practical impact on the present capital requirement function, as it is a closed-form expression.
- However, ES is turned down for other frameworks to keep consistency with the credit risk framework.
- This paper shows that the estimation precision of ES is not inferior to VaR, even at very high confidence levels. Parameter sensitivity and model validation is also on level with VaR.
  - As the arguments used against ES does not seem to hold, it could be worth considering the use of ES for credit risk regulation.

Paper II

- Osmundsen, Kjartan Kloster, Tore Selland Kleppe, and Atle Oglend (2019). **MCMC for Markov-switching models - Gibbs sampling vs. marginalized likelihood**.

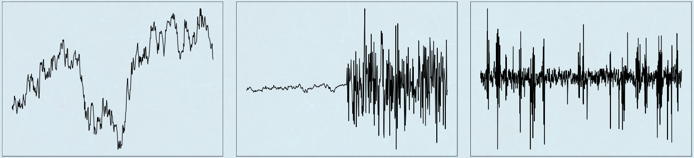


Motivation

- State-space models typically have a strong and non-linear dependence structure between the states and the parameters.
  - This can make it challenging to sample efficiently from their joint distribution.
- It is known that samplers that target the parameters directly (integrating out the states, thus reducing dimension) typically improves the performance of the resulting Markov chain.
- This paper tries to answer whether this also holds true for Markov-switching models.
  - A special case of state-space models, where the states only take discrete values.

Markov-switching models

- Mixture model governed by a finite state Markov chain, which is unobservable.
  - More flexible than standard parametric models



- We use a two-state Markov-switching vector autoregressive (MS-VAR) model as a basis for comparing different estimation methods (the methods discussed are not limited to the following model):

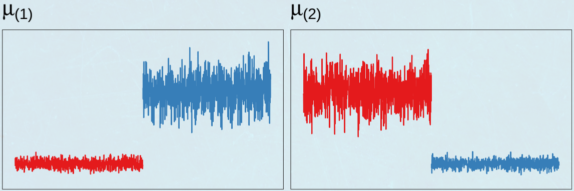
$$S_t = 1 \longrightarrow \boldsymbol{Y}_t = \phi_{(1)} \boldsymbol{Y}_{t-1} + \boldsymbol{\mu}_{(1)} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{(1)}\right).$$
$$S_t = 2 \longrightarrow \boldsymbol{Y}_t = \phi_{(2)} \boldsymbol{Y}_{t-1} + \boldsymbol{\mu}_{(2)} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{(2)}\right).$$

Sampling methods

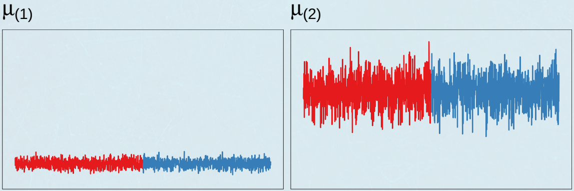
- To directly sample the parameters, we use the *forward algorithm* to compute the marginal log-likelihood analytically (by integrating over the latent states)
  - We use Hamiltonian Monte Carlo (HMC), which provides fast exploration of the relatively large parameter set. The no-U-turn extension eliminates the need for user-specified tuning parameters.
- As benchmark samplers, we use two different Gibbs samplers:
  - Direct Gibbs (DG), which sample each individual state S_t conditionally on the two closest states S_{t-1} and S_{t+1}.
  - Forward-backwards (FB) Gibbs sampler, which use recursive algorithms to sample the whole state vector in one operation.

Label-switching

- We use identical priors for each state, which makes the posterior likelihood invariant to relabelling of the states.
  - This can result in **label-switching** when you run multiple chains.



- This can be solved by using an identifiability constraint, usually by ordering on one of the parameters.



Simulation studies

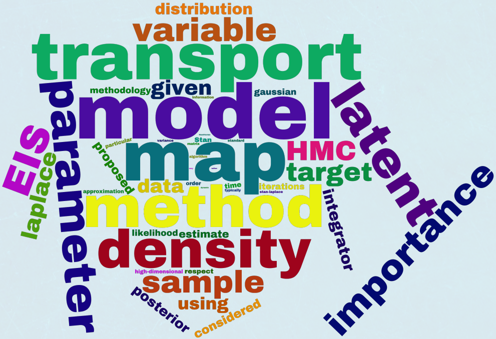
- Unrestricted two-state MS-VAR:
  $\boldsymbol{Y}_t = \phi_{(S_t)} \boldsymbol{Y}_{t-1} + \boldsymbol{\mu}_{(S_t)} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{(S_t)}\right)$
  - Simulated data, exchange rate data, interest rate data, crude oil data (each with 2 dimensions).
- Restricted two-state MS-VAR, with four autoregressive lags:
  $\boldsymbol{Y}_t = \phi_1 \boldsymbol{Y}_{t-1} + \phi_2 \boldsymbol{Y}_{t-2} + \phi_3 \boldsymbol{Y}_{t-3} + \phi_4 \boldsymbol{Y}_{t-4} + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{(S_t)}\right)$
  - US macro data set (3 dimensions):inflation, unemployment and interest rate.
- Unrestricted Markov-Switching Vector Error Correction (MS-VECM) model
  - Can be treated as an MS-VAR model with varying mean term.
  - The joint dynamics of natural gas and oil prices.

Paper II - Conclusion

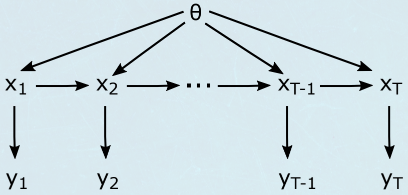
- The marginal method is found to be numerically robust, flexible with respect to model specification and easy to implement.
  - It mixes considerably better than the Gibbs samplers, and has approximately equal efficiency for all parameters.
- Implementations in suitable software packages results in similar computational times for both methods.
- Tailor-made implementations of the Gibbs samplers runs much faster, but require substantially greater coding efforts, and are not as easily adaptable to modelling changes or parameter restrictions.

Paper III

- Osmundsen, Kjartan Kloster, Tore Selland Kleppe, and Roman Liesenfeld (2019). **Importance Sampling-based Transport Map Hamiltonian Monte Carlo for Bayesian Hierarchical Models**.



Bayesian Hierarchical Models

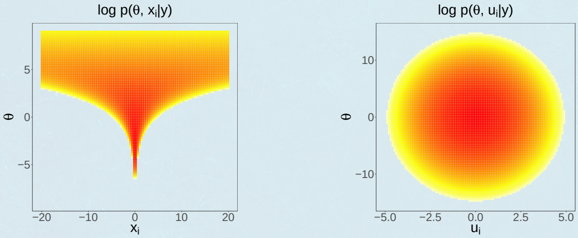


- State-space latent variable models
- We focus on models that are non-linear and/or non-Gaussian
  - Not feasible to analytically calculate the joint posterior for $(\boldsymbol{x}, \boldsymbol{\theta})$ or the marginal posterior for $\boldsymbol{\theta}$.

Motivation

- When the joint posterior for $(\boldsymbol{x}, \boldsymbol{\theta})$ have a complex dependence structure, HMC must be tuned for the most extremely scaled parts of the distribution to ensure full exploration, causing wasteful computations for the more moderately scaled parts.

- We propose a transport map method, which involves introducing a modified parametrization of the model, which is more well behaved for standard MCMC techniques.



Transport map

- We use a non-linear transport map, which relates the modified parametrization $\mathbf{q}'$ to the original parametrization $\mathbf{q}$. It is non-trivial for the latent variables only:

$$\mathbf{q} = \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{x} \end{bmatrix} = T(\mathbf{q}') = \begin{bmatrix} \boldsymbol{\theta} \\ \gamma_{\boldsymbol{\theta}}(\mathbf{u}) \end{bmatrix}, \ \mathbf{q}' = \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{u} \end{bmatrix},$$

- Applying the transport map to the joint posterior for $(\boldsymbol{x}, \boldsymbol{\theta})$, results in the following modified target distribution:

$$\tilde{\pi}(\boldsymbol{\theta}, \mathbf{u}|\mathbf{y}) \propto |\nabla_{\mathbf{u}} \gamma_{\boldsymbol{\theta}}(\mathbf{u})| \left[p(\boldsymbol{\theta}) p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x}|\boldsymbol{\theta}) \right]_{\mathbf{x}=\gamma_{\boldsymbol{\theta}}(\mathbf{u})}.$$

- After finding a suitable choice for $\gamma_{\boldsymbol{\theta}}(\mathbf{u})$, we simulate the modified target with standard HMC, and lastly apply the transport map to the resulting parameter chains.

Relation to importance sampling

- By letting $m(\mathbf{x}|\boldsymbol{\theta})$ denote the density of $\gamma_{\boldsymbol{\theta}}(\mathbf{u})$ when $\mathbf{u} \sim N(\mathbf{0}_D, \mathbf{I}_D)$, the modified target can also be expressed as

$$\tilde{\pi}(\boldsymbol{\theta}, \mathbf{u}|\mathbf{y}) \propto \mathcal{N}(\mathbf{u}|\mathbf{0}_D, \mathbf{I}_D) p(\boldsymbol{\theta}) \left[\frac{p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x}|\boldsymbol{\theta})}{m(\mathbf{x}|\boldsymbol{\theta})} \right]_{\mathbf{x}=\gamma_{\boldsymbol{\theta}}(\mathbf{u})}.$$

- The target reduces to $\mathcal{N}(\mathbf{u}|\mathbf{0}_D, \mathbf{I}_D) p(\boldsymbol{\theta}|\mathbf{y})$ if $m(\mathbf{x}|\boldsymbol{\theta}) = p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$
  - Creating a decoupling effect between $\boldsymbol{\theta}$ and $\mathbf{x}$.
  - Our strategy is to create a transport map that approximates this effect.

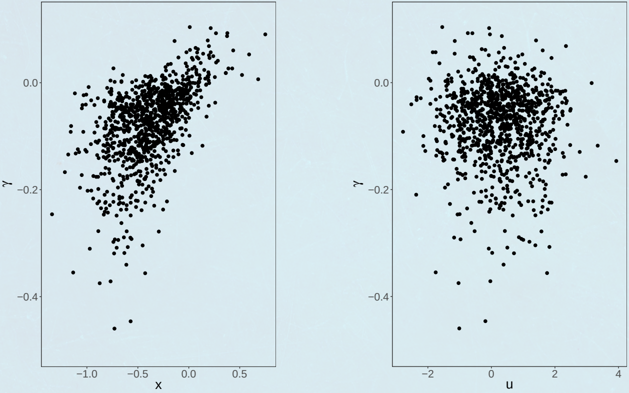
- Also, we recognize that the green part of the target is an importance weight targeting the marginal likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ when $\mathbf{u} \sim N(\mathbf{0}_D, \mathbf{I}_D)$
  - Our methodology may be seen as a special case of the Pseudo-marginal HMC of Lindsten and Doucet (2016), based on a single importance weight.

Simulation study

- HMC implementation:
  - Stan (Standard HMC with no-U-turn sampler)
  - HMC with specialized integrator by Lindsten and Doucet (LD)
- Approximation methods for the transport map:
  - Laplace approximation
  - Efficient importance sampling (EIS)
  - Benchmark method corresponding to $m(\mathbf{x}|\boldsymbol{\theta}) = p(\mathbf{x}|\boldsymbol{\theta})$
- Models, applied to financial time series:
  - Three different models with univariate states, exhibiting significantly different, and variable signal-to-noise ratios.
  - A high-dimensional application.

Stochastic volatility model

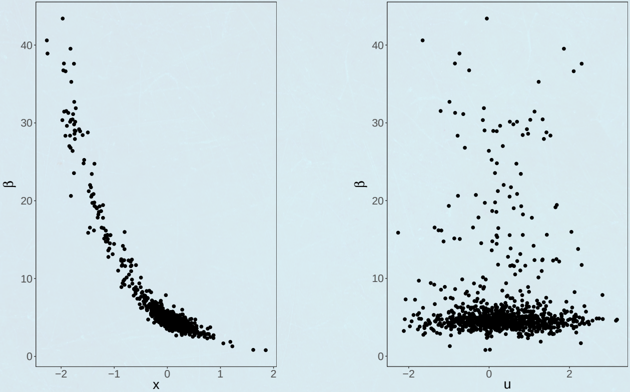
Low signal-to-noise ratio.



	LD-EIS	LD-Laplace	Stan-Laplace
ESS/s (relative to benchmark)	0.07	1.72	1.52

Gamma model for realized volatilities

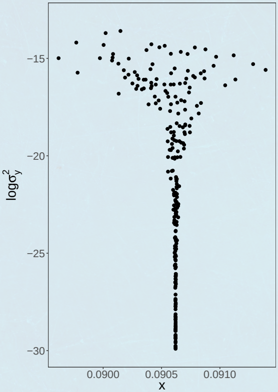
Medium signal-to-noise ratio.

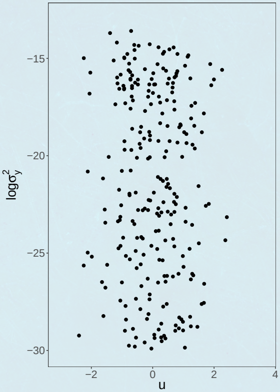


	LD-EIS	LD-Laplace	Stan-Laplace
ESS/s (relative to benchmark)	0.6	10.6	8.13

Constant elasticity of variance diffusion model

High signal-to-noise ratio.





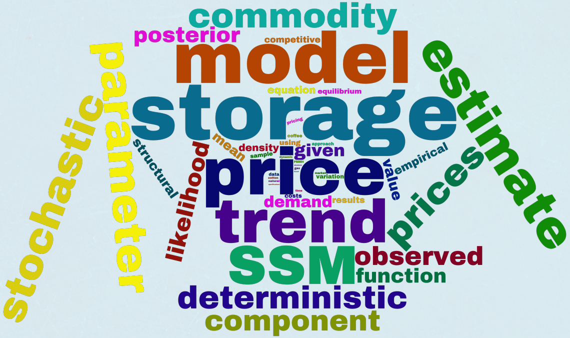
	Benchmark	LD-EIS	LD-Laplace	Stan-Laplace
ESS/s	-	1.6	16.3	1.8

Paper III - Conclusion

- For models with high-dimensional latent variables, the proposed methodology can lead to large speedups relative to relevant benchmarks, while still being easily implemented.
- Regarding the tradeoff between optimal accuracy and computational cost, the main insight of the paper is that rather crude transport maps are sufficient in this framework.
  - Contrary to most of the importance sampling literature, where typically very accurate importance densities are required.

Paper IV

- Osmundsen, Kjartan Kloster, Tore Selland Kleppe, Roman Liesenfeld, and Atle Oglend (2020). **Estimating the Competitive Storage Model with Stochastic Trends in Commodity Prices**.



Motivation

- Economic theories are often developed in a stationary context. This is also the case for the competitive storage model.
  – Commodity price data is often non-stationary. May lead to biased estimates for structural parameters, which determines important quantities like the price elasticity of demand and storage cost.
- Our paper propose a state-space model for commodity prices that extend the competitive storage model with a stochastic trend component.
  – Included to capture low-frequency price variations the storage model is unable to explain.

The competitive storage model

The model of Deaton and Laroque (1992) assumes:

- Independent and identically distributed shocks (supply/harvest)
- Costly storage
  - Commodity depreciation
  - Interest rate
- Non-negative storage
- A deterministic demand function $D(p_t)$, and its inverse function $P(x_t)$
- Speculators are assumed to hold rational expectations

Optimal storage policy implies $p_t = \max\left[P(x_t), \beta E_t p_{t+1}\right]$.

$P(x_t)$: The current price, given by supply and demand.
$\beta E_t p_{t+1}$: The expected price of next period, discounted for storage cost.

Pricing regimes

- We extend the model by including an upper limit of storage capacity, $C \geq 0$, leading to the following price function:

$$f(x) = \min\left[P(x - C), \max\left\{P(x), \beta E_t p_{t+1}\right\}\right]$$

- $P(x)$ is a deterministic function, while $\beta E_t p_{t+1}$ is solved numerically.
- The pricing function exhibits three different pricing regimes:
  - Stock-out regime, where $f(x) = P(x)$. Positive price spikes.
  - No-arbitrage regime, where $f(x) = \beta E_t p_{t+1}$. Low price volatility.
  - Full capacity regime, where $f(x) = P(x - C)$. Negative price spikes.

Stochastic trend

We start by expressing the storage model as a state-space model for observed log-prices p_t and latent stock process x_t:

$$p_t = \log f(x_t),$$
$$x_t = (1-\delta)\left[x_{t-1} - D\left(f(x_{t-1})\right)\right] + z_t, \qquad z_t \sim \text{iid } N(0,1),$$

Then, we add a stochastic trend:

$$p_t = k_t + \log f(x_t),$$
$$k_t = k_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \text{iid } N(0, v^2),$$

Objective: Estimate the storage model's structural parameters $\boldsymbol{\theta}$ for given price data, together with the latent parameters (**k** and **x**).

Bayesian inference: Particle Markov chain Monte Carlo

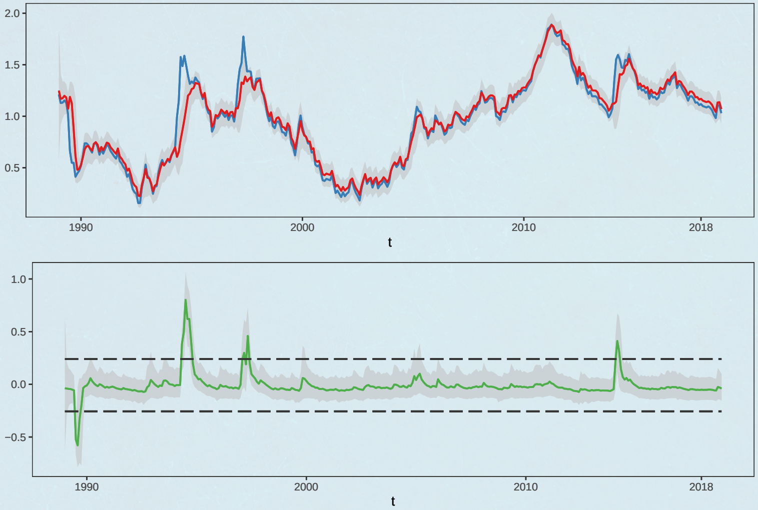
- The marginal likelihood for the observed prices is not available in closed form, as it is non-linear in the latent states.
  - Only a few approximation techniques are applicable, due to the discontinuous derivatives of the pricing function.
- Particle marginal Metropolis-Hastings approach (Andrieu et. al.,2010).
  - Uses unbiased MC estimates of the marginal likelihood inside a standard MH algorithm targeting the marginal posterior of the parameters.

Application

- The estimation methodology is applied to monthly commodity prices, for natural gas, coffee, cotton and aluminium.
- Compare the storage state-space model (SSM) to the restricted model that results in the absence of storage.
  - Adds significantly to explaining the observed commodity price behaviour.
- Compare the stochastic trend specification to parametric trends (linear trend and cubic spline trends).
  - The stochastic trend fits the price data much better than the deterministic trends.
- Residual diagnostics
  - Storage SSM outperforms the deterministic trend models in explaining the observed distributional properties of commodity prices.
  - The deterministic trends have a slight advantage in the approximation of the volatility dynamics.

Particle filter results: Coffee

Log price (blue), stochastic trend (red) and storage component (green).



Paper IV - Conclusion

- The paper proposes a SSM for commodity prices that combines the competitive storage model with a stochastic trend.
- Empirical application to four commodity markets shows that the stochastic trend SSM is favoured over deterministic trends.
- The stochastic trend SSM identifies structural parameters that differ from those obtained by deterministic trend specifications.
  - Substantially larger price elasticities of demand, for all four commodities.
  - Lower storage costs, except for natural gas.