Particle Methods in Econometrics

Robert Kohn
School of Economics, University of New South Wales
Joint work with Mike Pitt (Economics, Warwick), Marcel Scharth, UNSW, Minh Ngoc Tran, UNSW

Vienna, NOV 22 2013

Summary of the talk

- Motivating Example: Generalized Linear Mixed Model.

Summary of the talk

- Motivating Example: Generalized Linear Mixed Model.
- Bayesian inference when the likelihood is intractable but can be estimated unbiasedly.

Summary of the talk

- Motivating Example: Generalized Linear Mixed Model.
- Bayesian inference when the likelihood is intractable but can be estimated unbiasedly.
- Inference using the unbiased likelihood and either Importance Sampling Squared or MCMC on the parameters.

Summary of the talk

- Motivating Example: Generalized Linear Mixed Model.
- Bayesian inference when the likelihood is intractable but can be estimated unbiasedly.
- Inference using the unbiased likelihood and either Importance Sampling Squared or MCMC on the parameters.
- In either case we quantify an optimal precision for the estimator of the log likelihood.

Summary of the talk

- Motivating Example: Generalized Linear Mixed Model.
- Bayesian inference when the likelihood is intractable but can be estimated unbiasedly.
- Inference using the unbiased likelihood and either Importance Sampling Squared or MCMC on the parameters.
- In either case we quantify an optimal precision for the estimator of the log likelihood.
- We compare the the efficiency of the estimator based on the estimated likelihood against the corresponding scheme with a known likelihood.

Summary of the talk

- Motivating Example: Generalized Linear Mixed Model.
- Bayesian inference when the likelihood is intractable but can be estimated unbiasedly.
- Inference using the unbiased likelihood and either Importance Sampling Squared or MCMC on the parameters.
- In either case we quantify an optimal precision for the estimator of the log likelihood.
- We compare the the efficiency of the estimator based on the estimated likelihood against the corresponding scheme with a known likelihood.
- We make specific assumptions (which we can justify) on the error in the estimator of the log likelihood.

Generalised multinomial logit application; Utility Analysis

- The generalised multinomial logit (GMNL) model of Fiebig, Keane, Louviere , wasi (2010) specifies the utility of individual i from choosing alternative j at occasion t is

$$U_{ijt} = \beta_{0ij} + \sum_{k=1}^{K} \beta_{ki} x_{kijt} + \varepsilon_{ijt},$$

$$i = 1, \ldots, I \qquad j = 1, \ldots, J \qquad t = 1, \ldots, T,$$

where x_{kijt} are observed attributes for choice j, β_{ki} are heterogenous utility weights and ε_{ijt} are i.i.d. idiosyncratic errors following the extreme value distribution.

Choice Probabilities

- As in the standard multinomial logit model, the choice probability conditional on the observed attributes and utility weights have the simple closed form expression. i chooses j at time t,

$$\Pr(i,j,t|X_{it},\beta_i) = \frac{\exp(\beta_{0ij} + \sum_{k=1}^{K} \beta_{ki}x_{kijt})}{\sum_{h=1}^{J} \exp(\beta_{0ih} + \sum_{k=1}^{K} \beta_{ki}x_{kiht})}, \tag{1}$$

Choice Probabilities

- As in the standard multinomial logit model, the choice probability conditional on the observed attributes and utility weights have the simple closed form expression. i chooses j at time t,

$$\Pr(i,j,t|X_{it},\beta_i) = \frac{\exp(\beta_{0ij} + \sum_{k=1}^{K} \beta_{ki} x_{kijt})}{\sum_{h=1}^{J} \exp(\beta_{0ih} + \sum_{k=1}^{K} \beta_{ki} x_{kiht})}, \tag{1}$$

- The model for the utility weights is

$$\beta_{0ij} = \beta_{0j} + \eta_{0i}, \qquad \eta_{0i} \sim N(0,\sigma_0^2),$$
$$\beta_{ki} = \lambda_i \beta_k + \gamma \eta_{ki} + (1-\gamma)\lambda_i \eta_{ki}, \qquad \eta_{ki} \sim N(0,\sigma_k^2), \qquad k = 1,\ldots,$$
$$\lambda_i = \exp(-\delta/2 + \delta \zeta_i), \qquad \zeta_i \sim N(0,1),$$

Utility weights

- The model for the utility weights is

$$\beta_{0ij} = \beta_{0j} + \eta_{0i}, \qquad \eta_{0i} \sim N(0, \sigma_0^2),$$
$$\beta_{ki} = \lambda_i \beta_k + \gamma \eta_{ki} + (1-\gamma)\lambda_i \eta_{ki}, \qquad \eta_{ki} \sim N(0, \sigma_k^2), \qquad k = 1, \ldots,$$
$$\lambda_i = \exp(-\delta/2 + \delta \zeta_i), \qquad \zeta_i \sim N(0,1),$$

where β_{0j} are alternative specific constants (ASC) and λ_i are scaling coefficients. The parameter vector is
$\theta = (\beta_{01}, \ldots \beta_{0J}, \sigma_0^2, \beta_1, \ldots, \beta_K, \sigma_1^2, \ldots, \sigma_K^2, \delta^2, \gamma)'$.

Empirical Application

- We consider an empirical application to the pap smear data set used in the original paper by Fiebig et al.

Empirical Application

- We consider an empirical application to the pap smear data set used in the original paper by Fiebig et al.
- In this data set, $I = 79$ women choose whether or not to have a pap smear exam ($J = 2$) on $T = 32$ choice occasions.

Empirical Application

- We consider an empirical application to the pap smear data set used in the original paper by Fiebig et al.
- In this data set, $I = 79$ women choose whether or not to have a pap smear exam ($J = 2$) on $T = 32$ choice occasions.
- We let the observed choice for individual i at occasion t be $y_{it} = 1$ if the woman chooses to take the test and $y_{it} = 0$ otherwise.

Empirical Application

- We consider an empirical application to the pap smear data set used in the original paper by Fiebig et al.
- In this data set, $I = 79$ women choose whether or not to have a pap smear exam ($J = 2$) on $T = 32$ choice occasions.
- We let the observed choice for individual i at occasion t be $y_{it} = 1$ if the woman chooses to take the test and $y_{it} = 0$ otherwise.
- The next table lists the choice attributes and the associated coefficients.

Empirical Application

- We consider an empirical application to the pap smear data set used in the original paper by Fiebig et al.
- In this data set, $I = 79$ women choose whether or not to have a pap smear exam ($J = 2$) on $T = 32$ choice occasions.
- We let the observed choice for individual i at occasion t be $y_{it} = 1$ if the woman chooses to take the test and $y_{it} = 0$ otherwise.
- The next table lists the choice attributes and the associated coefficients.
- We normalise the utility of not taking the test to zero.

Choice attributes

Table : CHOICE ATTRIBUTES FOR THE PAP SMEAR DATA SET

.

Choice attributes	Values	Associated parameters
Constant for test	1	β_0, σ_0^2
Whether you know doctor	0 (no), 1 (yes)	β_1, σ_1^2
Whether doctor is male	0 (no), 1 (yes)	β_2, σ_2^2
Whether test is due	0 (no), 1 (yes)	β_3, σ_3^2
Whether doctor recommends	0 (no), 1 (yes)	β_4, σ_4^2
Test cost	$\{0, 10, 20, 30\}/10$	β_5

Bayesian Inference

- Posterior density

$$\pi(\theta) = p(\theta | y) = \frac{p(y | \theta)\, p(\theta)}{p(y)}$$

where $p(y | \theta)\, p(\theta)$ is known pointwise but $p(y)$ is not.

Bayesian Inference

- Posterior density

$$\pi(\theta) = p(\theta | y) = \frac{p(y|\theta)\,p(\theta)}{p(y)}$$

where $p(y|\theta)\,p(\theta)$ is known pointwise but $p(y)$ is not.

- Wish to estimate

$$\int h(\theta)\pi(\theta)d\theta$$

Bayesian Inference

- Posterior density

$$\pi(\theta) = p(\theta | y) = \frac{p(y|\theta)\, p(\theta)}{p(y)}$$

where $p(y|\theta)\, p(\theta)$ is known pointwise but $p(y)$ is not.

- Wish to estimate

$$\int h(\theta)\pi(\theta)d\theta$$

- MCMC have been used extensively to sample (approximately) from $\pi(\theta)$.

MCMC with Intractable Likelihood Function

- Consider now the scenario where $p(y|\theta)$ cannot be evaluated.

MCMC with Intractable Likelihood Function

- Consider now the scenario where $p\left(y \mid \theta\right)$ cannot be evaluated.
- For latent variable models

$$p\left(y \mid \theta\right) = \int p\left(x, y \mid \theta\right) dx$$

where $p\left(x, y \mid \theta\right)$ is often known but $p\left(y \mid \theta\right)$ is not.

MCMC with Intractable Likelihood Function

- Consider now the scenario where $p\left(y|\,\theta\right)$ cannot be evaluated.
- For latent variable models

$$p\left(y|\,\theta\right) = \int p\left(x, y|\,\theta\right) dx$$

where $p\left(x, y|\,\theta\right)$ is often known but $p\left(y|\,\theta\right)$ is not.

- More accurately, we know $p(y|x, \theta)$ and $p(x|\theta)$ and can generate from $p(x|\theta)$.

MCMC with Intractable Likelihood Function

- Consider now the scenario where $p(y|\theta)$ cannot be evaluated.
- For latent variable models

$$p(y|\theta) = \int p(x, y|\theta)\, dx$$

where $p(x, y|\theta)$ is often known but $p(y|\theta)$ is not.

- More accurately, we know $p(y|x, \theta)$ and $p(x|\theta)$ and can generate from $p(x|\theta)$.
- Standard MCMC approaches consists of sampling from

$$p(\theta, x|y) = \frac{p(x, y|\theta)\, p(\theta)}{p(y)}$$

by updating successively x and θ.

MCMC with Intractable Likelihood Function

- An alternative approach based on the availability of a non-negative unbiased likelihood estimator is possible.

MCMC with Intractable Likelihood Function

- An alternative approach based on the availability of a non-negative unbiased likelihood estimator is possible.
- In physics, first appeared in Lin, Liu & Sloan (2000). In statistics, Beaumont (2003), Andrieu, Berthelesen, D., Roberts (2006), Andrieu & Roberts (2009).

MCMC with Intractable Likelihood Function

- An alternative approach based on the availability of a non-negative unbiased likelihood estimator is possible.
- In physics, first appeared in Lin, Liu & Sloan (2000). In statistics, Beaumont (2003), Andrieu, Berthelesen, D., Roberts (2006), Andrieu & Roberts (2009).
- Particle Marginal Metropolis Hastings sampler for state-space models (Andrieu, D. & Holenstein, 2009, 2010).

- An alternative approach based on the availability of a non-negative unbiased likelihood estimator is possible.
- In physics, first appeared in Lin, Liu & Sloan (2000). In statistics, Beaumont (2003), Andrieu, Berthelesen, D., Roberts (2006), Andrieu & Roberts (2009).
- Particle Marginal Metropolis Hastings sampler for state-space models (Andrieu, D. & Holenstein, 2009, 2010).
- There is a nice paper by Andrieu and Vihola (2012) Convergence properties of pseudo-marginal Markov chain Monte Carlo that is related to our work.

MCMC with an Intractable Likelihood Function

- Denote by $\widehat{p}\left(y|\theta, u\right)$ the unbiased non-negative likelihood estimator function of the r.v. u of density $m\left(\left.u\right|\theta\right)$; i.e.

$$p\left(y|\theta\right) = \int \widehat{p}\left(y|\theta, u\right) m\left(\left.u\right|\theta\right) du$$

MCMC with an Intractable Likelihood Function

- Denote by $\widehat{p}\left(y|\theta, u\right)$ the unbiased non-negative likelihood estimator function of the r.v. u of density $m\left(\left.u\right|\theta\right)$; i.e.

$$p\left(y|\theta\right)=\int\widehat{p}\left(y|\theta,u\right)m\left(\left.u\right|\theta\right)du$$

- Given $\left(\theta, \widehat{p}\left(y|\theta, u\right)\right)$ then sample $\theta' \sim q\left(\left.\cdot\right|\theta\right)$, $u' \sim m\left(\left.\cdot\right|\theta'\right)$ and accept $\left(\theta', \widehat{p}\left(y|\theta', u'\right)\right)$ with a MH probability.

MCMC with an Intractable Likelihood Function

- Denote by $\widehat{p}(y|\theta, u)$ the unbiased non-negative likelihood estimator function of the r.v. u of density $m(u|\theta)$; i.e.

$$p(y|\theta) = \int \widehat{p}(y|\theta, u)\, m(u|\theta)\, du$$

- Given $(\theta, \widehat{p}(y|\theta, u))$ then sample $\theta' \sim q(\cdot|\theta)$, $u' \sim m(\cdot|\theta')$ and accept $(\theta', \widehat{p}(y|\theta', u'))$ with a MH probability.
- The MCMC has $p(\theta|y)$ as its marginal distribution whatever the variance of $\widehat{p}(y|\theta, u)$.

MCMC with Intractable Likelihood Function

- This algorithm is a M-H sampler targeting

$$\widehat{\pi}(\theta, u) \propto \widehat{p}(y|\theta, u)\, m(u|\theta)\, p(\theta)$$

using the proposal

$$q(\theta'|\theta)\, m(u'|\theta').$$

MCMC with Intractable Likelihood Function

- This algorithm is a M-H sampler targeting

$$\widehat{\pi}(\theta, u) \propto \widehat{p}(y|\theta, u)\, m(u|\theta)\, p(\theta)$$

using the proposal

$$q(\theta'|\theta)\, m(u'|\theta').$$

- Crucially unbiasedness provides that the marginal is:

$$\widehat{\pi}(\theta) = \pi(\theta) = p(\theta|y).$$

Importance Sampling Estimator

- Assume that

$$p\left(y\middle|\,\theta\right) = \int p\left(x, y\middle|\,\theta\right) dx.$$

Importance Sampling Estimator

- Assume that

$$p\left(y\right|\theta) = \int p\left(x, y\right|\theta)\, dx.$$

- Let $g(x|y,\theta)$ be an Importance Sampling (IS) density then

$$\widehat{p}(y|\theta, u) = \frac{1}{N}\sum_{k=1}^{N}\omega(x^k, \theta),$$

where the x^k are iid samples from $g(x|y;\theta)$, u is the vector of r.v. used to generate the x^k and

$$\omega(x, \theta) = \frac{p\left(x, y\right|\theta)}{g(x|y;\theta)}.$$

Importance Sampling Estimator

- Assume that

$$p\left(y\right|\theta) = \int p\left(x, y\right|\theta)\,dx.$$

- Let $g(x|y,\theta)$ be an Importance Sampling (IS) density then

$$\widehat{p}(y|\theta,u) = \frac{1}{N}\sum_{k=1}^{N}\omega(x^k,\theta),$$

where the x^k are iid samples from $g(x|y;\theta)$, u is the vector of r.v. used to generate the x^k and

$$\omega(x,\theta) = \frac{p\left(x, y\right|\theta)}{g(x|y;\theta)}.$$

- $\widehat{p}(y|\theta,u)$ is unbiased of variance inversely proportional to N.

Importance Sampling Estimator: Panel DATA

- Assume that

$$p(y|\theta) = \prod_{t=1}^{T} p(y_t|\theta)$$

$$p(y_t|\theta) = \int p(y_t|x_t;\theta)p(x_t|\theta)dx_t.$$

Importance Sampling Estimator: Panel DATA

- Assume that

$$p(y|\theta) = \prod_{t=1}^{T} p(y_t|\theta)$$
$$p(y_t|\theta) = \int p(y_t|x_t;\theta)p(x_t|\theta)dx_t.$$

- Let $g_t(x_t|y_t,\theta)$ be an Importance Sampling (IS) density. Then

$$\widehat{p}(y_t|\theta,u_t) = \frac{1}{N}\sum_{k=1}^{N}\omega(x_t^k,\theta),$$

where the x_t^k are iid samples from $g(x_t|y_t;\theta)$, u is the vector of r.v. used to generate the x_t^k and

$$\omega(x_t,\theta) = \frac{p(y_t|x_t;\theta)p(x_t;\theta)}{g(x_t|y_t;\theta)}.$$

Importance Sampling Estimator: Panel DATA II

- Assume that

$$\widehat{p}(y|\theta) = \prod_{t=1}^{T} \widehat{p}(y_t|\theta) = \prod_{t=1}^{T} \frac{1}{N} \sum_{k=1}^{N_t} \omega(x_t^k, \theta),$$

Importance Sampling Estimator: Panel DATA II

- Assume that

$$\widehat{p}(y|\theta) = \prod_{t=1}^{T} \widehat{p}(y_t|\theta) = \prod_{t=1}^{T} \frac{1}{N} \sum_{k=1}^{N_t} \omega(x_t^k, \theta),$$

- $\widehat{p}(y_t|\theta, u_t)$ is unbiased of variance inversely proportional to N_t.

Importance sampling squared. I

- Let $h(\theta)$ be a function of θ. We wish to estimate

$$\Delta(h) = \int h(\theta)p(\theta|y)d\theta$$
$$= I(h)/I(1)$$

Importance sampling squared. I

- Let $h(\theta)$ be a function of θ. We wish to estimate

$$\Delta(h) = \int h(\theta) p(\theta|y) d\theta$$
$$= I(h)/I(1)$$

- where

$$I(h) = \int h(\theta) p(y|\theta) p(\theta) d\theta$$

Importance sampling squared. I

- Let $h(\theta)$ be a function of θ. We wish to estimate

$$\Delta(h) = \int h(\theta) p(\theta|y) d\theta$$
$$= I(h)/I(1)$$

- where

$$I(h) = \int h(\theta) p(y|\theta) p(\theta) d\theta$$

- Define,

$$\widetilde{I}(h) = \int h(\theta) \widehat{p}(y|\theta, u) p(\theta) d\theta$$

Importance sampling squared. II

- Let $q(\theta)$ be an importance density.

$$\widetilde{I}(h) = \int h(\theta)\widehat{p}(y|\theta, u)p(\theta)d\theta \quad = \int h(\theta)\frac{\widehat{p}(y|\theta, u)p(\theta)}{q(\theta)}q(\theta)d\theta$$

Importance sampling squared. II

- Let $q(\theta)$ be an importance density.

$$\widetilde{I}(h) = \int h(\theta)\widehat{p}(y|\theta,u)p(\theta)d\theta \quad = \int h(\theta)\frac{\widehat{p}(y|\theta,u)p(\theta)}{q(\theta)}q(\theta)d\theta$$

- Then

$$\widehat{I}(h) = \frac{1}{M}\sum_{j=1}^{M} h(\theta^j)\frac{\widehat{p}(y|\theta^j,u)p(\theta^j)}{q(\theta^j)}$$

where $\theta^j \sim q(\theta)$, is the Importance squared estimator of $I(h)$.

Importance sampling squared. II

- Let $q(\theta)$ be an importance density.

$$\widetilde{I}(h) = \int h(\theta)\widehat{p}(y|\theta, u)p(\theta)d\theta \quad = \int h(\theta)\frac{\widehat{p}(y|\theta, u)p(\theta)}{q(\theta)}q(\theta)d\theta$$

- Then

$$\widehat{I}(h) = \frac{1}{M}\sum_{j=1}^{M} h(\theta^j)\frac{\widehat{p}(y|\theta^j, u)p(\theta^j)}{q(\theta^j)}$$

where $\theta^j \sim q(\theta)$, is the Importance squared estimator of $I(h)$.

-

$$\widehat{\Delta}(h) = \frac{\widehat{I}(h)}{\widehat{I}(1)}$$

- A state space model is a complex latent variable model.

- A state space model is a complex latent variable model.
-

$$p(y, x|\theta) = p(y|x;\theta)p(x|\theta)$$

$$p(y|x;\theta)\prod_{t=1}^{T} g(y_t|x_t;\theta)$$

$$p(x|\theta) = f(x_1|\theta)\prod_{t=2}^{T} f(x_t|x_{t-1};\theta)$$

SMC for state space models

- $$p(y|\theta) = p(y_1|\theta) \prod_{t=2}^{T} p(y_t|y_{t-1};\theta)$$

- $$p(y|\theta) = p(y_1|\theta) \prod_{t=2}^{T} p(y_t|y_{t-1};\theta)$$

- Omit θ for convenience.

$$p(y_t|y_{t-1}) = \int \left(\int w(x_t, x_{t-1}) g(x_t|x_{t-1}) dx_t \right) p(x_{t-1}|y_{1:t-1}) dx_{t-1}$$

$$w(x_t, x_t - 1) = \frac{p(y_t|x_t)p(x_t|x_{t-1})}{g(x_t|x_{t-1})}$$

SMC II

- If we "know" $p(x_{t-1}|y_{1:t-1})$ and have samples $x_{t-1}^j, j = 1, \ldots, M$ from it, then we can generate x_t^j from $g(x_t|x_{t-1})$ and

$$\widehat{p}(y_t|y_{t-1}) = \frac{1}{M}\sum_{j=1}^{M} w(x_t^j, x_{t-1}^j)$$

SMC II

- If we "know" $p(x_{t-1}|y_{1:t-1})$ and have samples $x_{t-1}^j, j = 1, \ldots, M$ from it, then we can generate x_t^j from $g(x_t|x_{t-1})$ and

$$\widehat{p}(y_t|y_{t-1}) = \frac{1}{M} \sum_{j=1}^{M} w(x_t^j, x_{t-1}^j)$$

-

$$\begin{aligned}
\widehat{p}(y|\theta) &= \prod_{t=1}^{T} \widehat{p}(y_t|y_{t-1};\theta) \\
&= \prod_{t=1}^{T} \frac{1}{M} \sum_{j=1}^{M} w(x_t^j, x_{t-1}^j;\theta)
\end{aligned}$$

SMC III

- $$\widehat{p}(y|\theta) = \prod_{t=1}^{T} \widehat{p}(y_t|y_{t-1};\theta)$$
  $$= \prod_{t=1}^{T} \frac{1}{M} \sum_{j=1}^{M} w(x_t^j, x_{t-1}^j;\theta)$$

SMC III

- $$\begin{aligned} \widehat{p}(y|\theta) &= \prod_{t=1}^{T} \widehat{p}(y_t|y_{t-1};\theta) \\ &= \prod_{t=1}^{T} \frac{1}{M} \sum_{j=1}^{M} w(x_t^j, x_{t-1}^j;\theta) \end{aligned}$$
- Note that $\widehat{p}(y|\theta)$ is again unbiased. So SMC is another example of estimating a likelihood unbiasedly.

Inference for Non-linear Models

- We consider non-linear state space models (West and Harrison, Harvey). A classic highly non-linear model from Kitagawa (1996),

$$y_t = \frac{1}{20} x_t^2 + w_t, \quad w_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma_W^2\right)$$

$$x_t = \frac{1}{2} x_{t-1} + 25 \frac{x_{t-1}}{1 + x_{t-1}^2} + 8\cos(1.2t) + v_t, \quad v_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma_V^2\right),$$

  We follow Andrieu 2009 in having an initial distribution $x_1 \sim N(0, 5)$ and taking $\sigma_V^2 = 10$, and $\sigma_W^2 = 10$, with $T = 200$.

Inference for Non-linear Models

- We consider non-linear state space models (West and Harrison, Harvey). A classic highly non-linear model from Kitagawa (1996),

$$y_t = \frac{1}{20} x_t^2 + w_t, \quad w_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma_W^2\right)$$

$$x_t = \frac{1}{2} x_{t-1} + 25 \frac{x_{t-1}}{1 + x_{t-1}^2} + 8\cos(1.2t) + v_t, \quad v_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma_V^2\right),$$

  We follow Andrieu 2009 in having an initial distribution $x_1 \sim N(0, 5)$ and taking $\sigma_V^2 = 10$, and $\sigma_W^2 = 10$, with $T = 200$.
- Difficult/Expensive to perform standard MCMC.

Inference for Non-linear Models

- We consider non-linear state space models (West and Harrison, Harvey). A classic highly non-linear model from Kitagawa (1996),

$$y_t = \frac{1}{20} x_t^2 + w_t, \quad w_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma_W^2\right)$$

$$x_t = \frac{1}{2} x_{t-1} + 25 \frac{x_{t-1}}{1 + x_{t-1}^2} + 8 \cos(1.2t) + v_t, \quad v_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma_V^2\right),$$

  We follow Andrieu 2009 in having an initial distribution $x_1 \sim N(0, 5)$ and taking $\sigma_V^2 = 10$, and $\sigma_W^2 = 10$, with $T = 200$.
- Difficult/Expensive to perform standard MCMC.
- We sample from $p\left(\theta \mid y_{1:T}\right)$ using a Metropolis-Hastings sampler where $p\left(y_{1:T} \mid \theta\right)$ is estimated unbiasedly using a particle filter. We vary N and use random walk proposals for $\log \sigma_V$, $\log \sigma_W$. We use 100,000 MCMC steps.

Autocorrelation plots of parameters for Kitagawa model

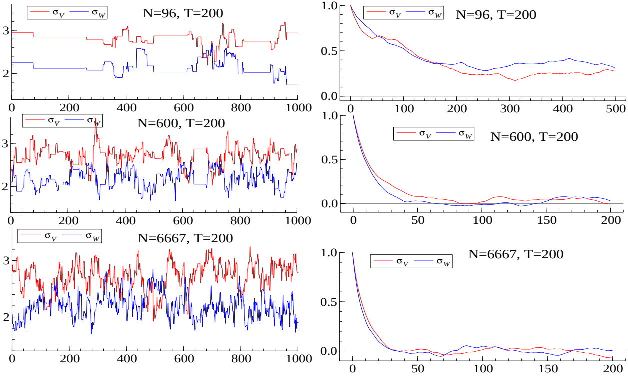


Figure : Autocorrelation of σ_V and σ_W of the MH sampler for various N in the PF

How to Select the Number of Samples

- A key issue from a practical point of view is how to select N?

How to Select the Number of Samples

- A key issue from a practical point of view is how to select N?
- If N is too small, then the algorithm mixes poorly and will require many MCMC iterations.

How to Select the Number of Samples

- A key issue from a practical point of view is how to select N?
- If N is too small, then the algorithm mixes poorly and will require many MCMC iterations.
- If N is too large, then each MCMC iteration or IS step is expensive.

How to Select the Number of Samples

- A key issue from a practical point of view is how to select N?
- If N is too small, then the algorithm mixes poorly and will require many MCMC iterations.
- If N is too large, then each MCMC iteration or IS step is expensive.
- Aim: We would like to provide guidelines on how to select N

MCMC with Intractable Likelihood Function

- Let $z = \log \widehat{p}_N\left(y|\theta, u\right) - \log p(y|\theta)$ be the error in the log-likelihood.

MCMC with Intractable Likelihood Function

- Let $z = \log \widehat{p}_N(y|\theta, u) - \log p(y|\theta)$ be the error in the log-likelihood.
- The proposal from which z arises is denoted $g_N(z|\theta)$.

- Let $z = \log \widehat{p}_N\left(y|\theta, u\right) - \log p(y|\theta)$ be the error in the log-likelihood.
- The proposal from which z arises is denoted $g_N(z|\theta)$.
- We can rewrite the extended target

$$\widehat{\pi}_N(\theta, z) = \pi(\theta)\exp(z)g_N(z|\theta)$$

which is directly related to $\widehat{\pi}_N(\theta, u)$ through the many-to-one transformation from u to z.

Inefficiency Measure

- We wish to estimate

$$\mu_h = \mathbb{E}_\pi \left[h(\theta) \right] \quad \text{by} \quad \widehat{\mu}_{h,n} = n^{-1} \sum_{j=1}^{n} h(\theta_j).$$

Then the IACT or inefficiency of the Markov chain IF_h is given by

$$IF_h = \frac{V_\pi(\widehat{\mu}_{h,n})}{V_\pi(h)/n}$$

Inefficiency Measure

- We wish to estimate

$$\mu_h = \mathbb{E}_\pi [h(\theta)] \quad \text{by} \quad \widehat{\mu}_{h,n} = n^{-1} \sum_{j=1}^{n} h(\theta_j).$$

Then the IACT or inefficiency of the Markov chain IF_h is given by

$$IF_h = \frac{V_\pi(\widehat{\mu}_{h,n})}{V_\pi(h)/n}$$

- The IACT, IF_h, quantifies the factor by which we need to increase the number of samples from the Markov chain relative to using iid samples from $\pi(\theta)$ to achieve a given precision.

Making Assumptions to Move Forward

Let $z = \log \widehat{p}_N(y|\theta, u) - \log p(y|\theta)$ be the error in the estimator of the log likelihood.

Assumptions.

- We assume that z is normally distributed. This implies that the "prior" density of z is

$$g_N(z|\theta) = \phi\left(z; -\gamma^2(\theta)/2N, \gamma^2(\theta)/N\right)$$

and the "posterior" density is

$$\pi_N(z|\theta) = \exp(z)g_N(z|\theta) = \phi\left(z; \gamma^2(\theta)/2N, \gamma^2(\theta)/N\right)$$

where $\phi(z; a, b^2)$ is a univariate normal of mean a, variance b^2.

Making Assumptions to Move Forward

Let $z = \log \widehat{p}_N(y|\theta, u) - \log p(y|\theta)$ be the error in the estimator of the log likelihood.

Assumptions.

- We assume that z is normally distributed. This implies that the "prior" density of z is

$$g_N(z|\theta) = \phi\left(z; -\gamma^2(\theta)/2N, \gamma^2(\theta)/N\right)$$

  and the "posterior" density is

$$\pi_N(z|\theta) = \exp(z) g_N(z|\theta) = \phi\left(z; \gamma^2(\theta)/2N, \gamma^2(\theta)/N\right)$$

  where $\phi(z; a, b^2)$ is a univariate normal of mean a, variance b^2.
- For a given value of σ^2 we set $N = N_{\sigma^2}(\theta) = \gamma(\theta)^2/\sigma^2$.

Consequences of the Assumptions

Under these assumptions,

- Both $g_N(z|\theta)$ and $\pi_N(z|\theta)$ are functions of σ^2 only and we write $g_N(z|\theta)$ and $\pi_N(z|\theta)$ as

$$g(z|\sigma^2) = \phi(z; -\sigma^2/2, \sigma^2), \quad \pi(z|\sigma^2) = \phi(z; \sigma^2/2, \sigma^2).$$

Consequences of the Assumptions

Under these assumptions,

- Both $g_N(z|\theta)$ and $\pi_N(z|\theta)$ are functions of σ^2 only and we write $g_N(z|\theta)$ and $\pi_N(z|\theta)$ as

$$g(z|\sigma^2) = \phi(z; -\sigma^2/2, \sigma^2), \quad \pi(z|\sigma^2) = \phi(z; \sigma^2/2, \sigma^2).$$

- θ and z are independent under $\widehat{\pi}_N(\theta, z)$.

Consequences of the Assumptions

Under these assumptions,

- Both $g_N(z|\theta)$ and $\pi_N(z|\theta)$ are functions of σ^2 only and we write $g_N(z|\theta)$ and $\pi_N(z|\theta)$ as

$$g(z|\sigma^2) = \phi(z; -\sigma^2/2, \sigma^2), \quad \pi(z|\sigma^2) = \phi(z; \sigma^2/2, \sigma^2).$$

- θ and z are independent under $\widehat{\pi}_N(\theta, z)$.
- So everything just depends on σ, which is the variance of Z, i.e., the variance of the log likelihood estimator.

Main Result: Computing Time

- We would like to choose σ minimize computing time for a give level of precision or inefficiency.

Main Result: Computing Time

- We would like to choose σ minimize computing time for a give level of precision or inefficiency.
- The optimal value of σ is about 1.

Main Result: Computing Time

- We would like to choose σ minimize computing time for a give level of precision or inefficiency.
- The optimal value of σ is about 1.
- If we make σ much bigger than 1 than we get really high computing time for a given inefficiency.

Main Result: Computing Time

- We would like to choose σ minimize computing time for a give level of precision or inefficiency.
- The optimal value of σ is about 1.
- If we make σ much bigger than 1 than we get really high computing time for a given inefficiency.
- If we make σ very small, i.e., very high number of particles, then we waste N.

Main Result: Computing Time

- We would like to choose σ minimize computing time for a give level of precision or inefficiency.
- The optimal value of σ is about 1.
- If we make σ much bigger than 1 than we get really high computing time for a given inefficiency.
- If we make σ very small, i.e., very high number of particles, then we waste N.
- If the proposal for θ is very good, then we want σ smaller. If the proposal is not very good, optimal σ will be larger.

Relative Upper Bounds on Inefficiency and Computing Time

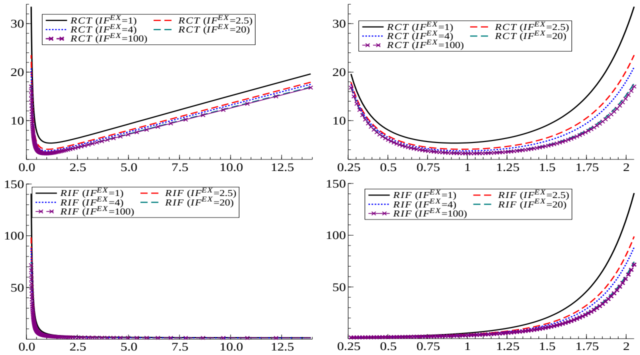


Figure : RCT_h^{U} (top) and RIF_h^{U} (bottom) against $1/\sigma^2$ (left) and σ (right). Different values of IF_h^{EX} are shown on each plot.

Empirical vs Asymptotic Distribution of Log-Likelihood Estimator

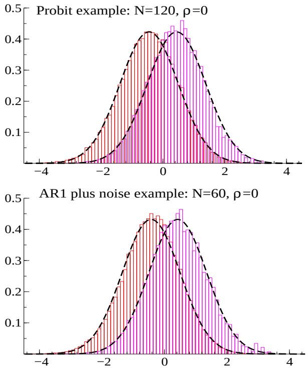


Figure : Histograms of proposed (red) and accepted (pink) values of z in PMCMC scheme. Overlayed are Gaussian pdfs from our simplifying Assumption for a target of $\sigma = 0.92$.

Importance sampling Squared

- Under the same assumptions as before, let $V_{IS}(\phi)$ be the variance of the IS estimator assuming that we use the exact likelihood for a given σ^2 (variance of the log likelihood).

Importance sampling Squared

- Under the same assumptions as before, let $V_{IS}(\phi)$ be the variance of the IS estimator assuming that we use the exact likelihood for a given σ^2 (variance of the log likelihood).
- Let $V_{IS^2}(\phi, \sigma^2)$ be the variance of the IS squared estimator for a given σ^2 (variance of the log likelihood estimator).

Importance sampling Squared

- Under the same assumptions as before, let $V_{IS}(\phi)$ be the variance of the IS estimator assuming that we use the exact likelihood for a given σ^2 (variance of the log likelihood).
- Let $V_{IS^2}(\phi, \sigma^2)$ be the variance of the IS squared estimator for a given σ^2 (variance of the log likelihood estimator).
- Then, define the inefficiency of IS-squared.

$$
\begin{aligned}
IF_{IS^2}(\sigma^2) &= \frac{V_{IS^2}(\phi, \sigma^2)}{V_{IS}(\phi)} \\
&= \exp(\sigma^2)
\end{aligned}
$$

Importance sampling Squared

- Under the same assumptions as before, let $V_{IS}(\phi)$ be the variance of the IS estimator assuming that we use the exact likelihood for a given σ^2 (variance of the log likelihood).
- Let $V_{IS^2}(\phi, \sigma^2)$ be the variance of the IS squared estimator for a given σ^2 (variance of the log likelihood estimator).
- Then, define the inefficiency of IS-squared.

$$IF_{IS^2}(\sigma^2) = \frac{V_{IS^2}(\phi, \sigma^2)}{V_{IS}(\phi)} = \exp(\sigma^2)$$

- Define Computing Time

$$CT_{IS^2} = \frac{IF_{IS^2}(\sigma^2)}{\sigma^2}$$

Importance sampling Squared

- Under the same assumptions as before, let $V_{IS}(\phi)$ be the variance of the IS estimator assuming that we use the exact likelihood for a given σ^2 (variance of the log likelihood).
- Let $V_{IS^2}(\phi, \sigma^2)$ be the variance of the IS squared estimator for a given σ^2 (variance of the log likelihood estimator).
- Then, define the inefficiency of IS-squared.

$$
\begin{aligned}
IF_{IS^2}(\sigma^2) &= \frac{V_{IS^2}(\phi, \sigma^2)}{V_{IS}(\phi)} \\
&= \exp(\sigma^2)
\end{aligned}
$$

- Define Computing Time

$$
CT_{IS^2} = \frac{IF_{IS^2}(\sigma^2)}{\sigma^2}
$$

- Optimum at $\sigma^2 = 1$.

Importance Sampling Squared II

- Above we assume that there is no overhead in obtaining the important sampling squared estimator and targeting the variance to be 1. Thus cost is proportional to the number of particles N.

Importance Sampling Squared II

- Above we assume that there is no overhead in obtaining the important sampling squared estimator and targeting the variance to be 1. Thus cost is proportional to the number of particles N.
- If there is also a fixed cost then that must also be taken into account.

Importance Sampling Squared II

- Above we assume that there is no overhead in obtaining the important sampling squared estimator and targeting the variance to be 1. Thus cost is proportional to the number of particles N.
- If there is also a fixed cost then that must also be taken into account.
- Let $\sigma^2(\theta) = \gamma(\theta)^2/N$ be the variance of the log likelihood.

Importance Sampling Squared II

- Above we assume that there is no overhead in obtaining the important sampling squared estimator and targeting the variance to be 1. Thus cost is proportional to the number of particles N.
- If there is also a fixed cost then that must also be taken into account.
- Let $\sigma^2(\theta) = \gamma(\theta)^2/N$ be the variance of the log likelihood.
- The computing time for IS^2 is

$$CT_{IS^2} = \exp(\gamma^2(\theta)/N)(\tau_1 + \tau_2 N)$$

which is minimized at $N^{opt}(\theta)$.

Generalised multinomial logit application; Utility Analysis

- The generalised multinomial logit (GMNL) model of Fiebig, Keane, Louviere , wasi (2010) specifies the utility of individual i from choosing alternative j at occasion t is

$$U_{ijt} = \beta_{0ij} + \sum_{k=1}^{K} \beta_{ki} x_{kijt} + \varepsilon_{ijt},$$

$$i = 1, \ldots, I \qquad j = 1, \ldots, J \qquad t = 1, \ldots, T,$$

where x_{kijt} are observed attributes for choice j, β_{ki} are heterogenous utility weights and ε_{ijt} are i.i.d. idiosyncratic errors following the extreme value distribution.

Choice Probabilities

- As in the standard multinomial logit model, the choice probability conditional on the observed attributes and utility weights have the simple closed form expression. i chooses j at time t,

$$\Pr(i, j, t | X_{it}, \beta_i) = \frac{\exp(\beta_{0ij} + \sum_{k=1}^{K} \beta_{ki} x_{kijt})}{\sum_{h=1}^{J} \exp(\beta_{0ih} + \sum_{k=1}^{K} \beta_{ki} x_{kiht})}, \quad (2)$$

Choice Probabilities

- As in the standard multinomial logit model, the choice probability conditional on the observed attributes and utility weights have the simple closed form expression. i chooses j at time t,

$$\Pr(i,j,t|X_{it},\beta_i) = \frac{\exp(\beta_{0ij} + \sum_{k=1}^{K} \beta_{ki} x_{kijt})}{\sum_{h=1}^{J} \exp(\beta_{0ih} + \sum_{k=1}^{K} \beta_{ki} x_{kiht})}, \tag{2}$$

- The model for the utility weights is

$$\begin{aligned}
\beta_{0ij} &= \beta_{0j} + \eta_{0i}, \qquad \eta_{0i} \sim N(0,\sigma_0^2), \\
\beta_{ki} &= \lambda_i \beta_k + \gamma \eta_{ki} + (1-\gamma)\lambda_i \eta_{ki}, \qquad \eta_{ki} \sim N(0,\sigma_k^2), \qquad k = 1, \ldots, \\
\lambda_i &= \exp(-\delta/2 + \delta \zeta_i), \qquad \zeta_i \sim N(0,1),
\end{aligned}$$

where β_{0j} are alternative specific constants (ASC) and λ_i are scaling coefficients. The parameter vector is
$\theta = (\beta_{01}, \ldots \beta_{0J}, \sigma_0^2, \beta_1, \ldots, \beta_K, \sigma_1^2, \ldots, \sigma_K^2, \delta^2, \gamma)'$.

Empirical Application

We consider an empirical application to the pap smear data set used in the original paper by Fiebig et al. In this data set, $I = 79$ women choose whether or not to have a pap smear exam ($J = 2$) on $T = 32$ choice occasions. We let the observed choice for individual i at occasion t be $y_{it} = 1$ if the woman chooses to take the test and $y_{it} = 0$ otherwise. The next table lists the choice attributes and the associated coefficients. We impose the restriction that $\sigma_5^2 = 0$ in our illustration since we have not found evidence of heterogeneity for this attribute beyond the scaling effect. We normalise the utility of not taking the test to zero.

Choice attributes

Table : Choice attributes for the pap smear data set

.

Choice attributes	Values	Associated parameters
Constant for test	1	β_0, σ_0^2
Whether you know doctor	0 (no), 1 (yes)	β_1, σ_1^2
Whether doctor is male	0 (no), 1 (yes)	β_2, σ_2^2
Whether test is due	0 (no), 1 (yes)	β_3, σ_3^2
Whether doctor recommends	0 (no), 1 (yes)	β_4, σ_4^2
Test cost	$\{0, 10, 20, 30\}/10$	β_5

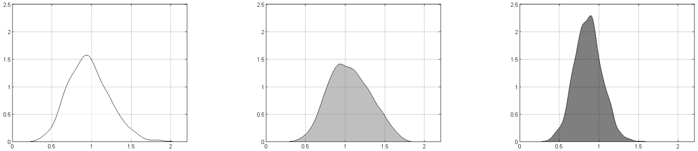
Likelihood Evaluation

Table : Generalized multinomial logit - log-likelihood evaluation for the parameters sampled from the mixture of multivariate t proposal.

The table shows the average variance, skewness and kurtosis of log-likelihood estimates 1,000 several draws from the importance density for the model parameters. The JB rejections row report the proportion of replications in which the Jarque-Bera tests rejects the null hypothesis of normality of the log-likelihood estimates at the 5% level.

	$N = 10,000$	$N = 20,000$	$\sigma^2 \approx 1$
Variance	1.661	0.856	1.024
Relative Var.	1.940	1.000	1.197
Skewness	0.008	0.001	-0.038
Kurtosis	2.955	2.972	3.003
JB rejections (5%)	0.059	0.059	0.055
Time (sec)	1.377	2.836	2.070

Distribution of log likelihood standard deviation



(a) Targeting a standard deviation of 1 on average

(b) Adapting the number of importance samples to target a log-likelihood standard deviation of 1 for each parameter

(c) Adapting the number of importance samples to target the optimal log-likelihood standard deviation for each parameter

Figure : Distribution of the log-likelihood standard deviation across different draws of the importance density for the parameters and different schemes to select the number of importance samples for estimating the likelihood.

Comparing different implementations

Table : Generalised multinomial logit - relative time normalised variances for posterior inference.

The table shows the relative variances for IS^2 for different numbers of importance samples u for estimating the likelihood.

	N=1,000	N=2,000	N=3,000	N=4,000	N_θ $(\sigma \approx 1)$	N_θ (optima
β_0	1.234	1.000	0.746	0.841	0.890	0.572
β_1	1.132	1.000	0.881	0.787	0.704	0.620
β_2	0.991	1.000	0.768	0.753	0.688	0.711
β_3	0.776	1.000	0.827	0.692	0.590	0.640
β_4	0.937	1.000	0.840	0.764	0.684	0.679
β_5	1.099	1.000	0.762	0.740	0.766	0.673
σ_0	1.225	1.000	0.679	0.729	0.814	0.679
σ_1	0.776	1.000	1.735	1.356	0.553	0.483
σ_2	1.192	1.000	0.900	0.816	0.615	0.919
σ_3	0.996	1.000	0.748	0.697	0.580	0.655
σ_4	2.080	1.000	0.974	0.841	0.815	0.871
Average	1.120	1.000	0.864	0.798	0.708	0.682

Discussion

- We have provided an approximate analysis of MCMC using unbiased likelihood estimator.

Discussion

- We have provided an approximate analysis of MCMC using unbiased likelihood estimator.
- We have placed very weak assumptions on the underlying (known likelihood) chain (reversible)).

Discussion

- We have provided an approximate analysis of MCMC using unbiased likelihood estimator.
- We have placed very weak assumptions on the underlying (known likelihood) chain (reversible)).
- The (asym) assumptions on the estimator error appears reasonable.

Discussion

- We have provided an approximate analysis of MCMC using unbiased likelihood estimator.
- We have placed very weak assumptions on the underlying (known likelihood) chain (reversible)).
- The (asym) assumptions on the estimator error appears reasonable.
- For a general proposal and under simplifying assumptions on the likelihood estimator, we can get guidelines on how to select σ: *as long as σ is around 1 then you are fine.*

Nonlinear structural models

- In general, an economic model with optimising agents and rational expectations can be written in the form

$$\mathbb{E}_t F\left(c_{t+1}, c_t, k_t, k_{t-1}, z_t, z_{t-1}, u_t \mid \theta\right) = 0 \tag{3}$$

where $\mathbb{E}_t$ denotes an expectation conditional on date t information;

Nonlinear structural models

- In general, an economic model with optimising agents and rational expectations can be written in the form

$$\mathbb{E}_t F\left(c_{t+1}, c_t, k_t, k_{t-1}, z_t, z_{t-1}, u_t \mid \theta\right) = 0 \tag{3}$$

  where $\mathbb{E}_t$ denotes an expectation conditional on date t information;
- c_t is a vector of choice variables (including forward-looking variables and jump variables);

Nonlinear structural models

- In general, an economic model with optimising agents and rational expectations can be written in the form

$$\mathbb{E}_t F\left(c_{t+1}, c_t, k_t, k_{t-1}, z_t, z_{t-1}, u_t \mid \theta\right) = 0 \qquad (3)$$

where $\mathbb{E}_t$ denotes an expectation conditional on date t information;

- c_t is a vector of choice variables (including forward-looking variables and jump variables);
- k_t is a vector of endogenous predetermined variables,

Nonlinear structural models

- In general, an economic model with optimising agents and rational expectations can be written in the form

$$\mathbb{E}_t F\left(c_{t+1}, c_t, k_t, k_{t-1}, z_t, z_{t-1}, u_t \mid \theta\right) = 0 \tag{3}$$

  where $\mathbb{E}_t$ denotes an expectation conditional on date t information;
- c_t is a vector of choice variables (including forward-looking variables and jump variables);
- k_t is a vector of endogenous predetermined variables,
- z_{t-1} is a vector of exogenous forcing variables, and u_t is a vector of independently and identically distributed (iid) shocks.

Nonlinear structural models

- In general, an economic model with optimising agents and rational expectations can be written in the form

$$\mathbb{E}_t F\left(c_{t+1}, c_t, k_t, k_{t-1}, z_t, z_{t-1}, u_t \mid \theta\right) = 0 \tag{3}$$

  where $\mathbb{E}_t$ denotes an expectation conditional on date t information;
- c_t is a vector of choice variables (including forward-looking variables and jump variables);
- k_t is a vector of endogenous predetermined variables,
- z_{t-1} is a vector of exogenous forcing variables, and u_t is a vector of independently and identically distributed (iid) shocks.
- θ is a vector of parameters

State space form

- Let $x_t = (c_t, k_t, z_t)$, and y_t the vector of observable variables in period t;

State space form

- Let $x_t = (c_t, k_t, z_t)$, and y_t the vector of observable variables in period t;
- we use a solution of the form

$$\begin{aligned} y_t &= Zx_t + \eta_t \\ x_t &= h(x_{t-1}, u_t) \end{aligned} \tag{4}$$

for some function $h(\cdot)$, where Z is a selection matrix of ones and zeros and η_t is observational noise.

Second Order Approximation

- Taking a second-order approximation, e.g. schmitt and grohe 2004, gives a system in the form

Second Order Approximation

- Taking a second-order approximation, e.g. schmitt and grohe 2004, gives a system in the form
-

$$x_t = d + Ex_{t-1} + Fu_t + \left(I \otimes x'_{t-1}\right) Gx_{t-1} + \left(I \otimes x'_{t-1}\right) Hu_t + \left(I \otimes u'_t\right) Ju_t.$$

E and F are the coefficient matrices from the first-order approximation of the model,

Second Order Approximation

- Taking a second-order approximation, e.g. schmitt and grohe 2004, gives a system in the form
- $$x_t = d + Ex_{t-1} + Fu_t + \left(I \otimes x'_{t-1}\right) Gx_{t-1} + \left(I \otimes x'_{t-1}\right) Hu_t + \left(I \otimes u'_t\right) Ju_t.$$

  E and F are the coefficient matrices from the first-order approximation of the model,
- d is a correction for volatility that does not appear in the first-order approximation

Second Order Approximation

- Taking a second-order approximation, e.g. schmitt and grohe 2004, gives a system in the form
-

$$x_t = d + Ex_{t-1} + Fu_t + \left(I \otimes x'_{t-1}\right) Gx_{t-1} \\ + \left(I \otimes x'_{t-1}\right) Hu_t + \left(I \otimes u'_t\right) Ju_t.$$

  E and F are the coefficient matrices from the first-order approximation of the model,
- d is a correction for volatility that does not appear in the first-order approximation
- The matrix G contains coefficients on the squares and cross-products of the model's endogenous variables;

Second Order Approximation

- Taking a second-order approximation, e.g. schmitt and grohe 2004, gives a system in the form
-
$$x_t = d + Ex_{t-1} + Fu_t + \left(I \otimes x'_{t-1}\right) Gx_{t-1} + \left(I \otimes x'_{t-1}\right) Hu_t + \left(I \otimes u'_t\right) Ju_t.$$

  E and F are the coefficient matrices from the first-order approximation of the model,
- d is a correction for volatility that does not appear in the first-order approximation
- The matrix G contains coefficients on the squares and cross-products of the model's endogenous variables;
- the matrix H has coefficients for interaction terms between lagged endogenous variables and current-period shocks;

Second Order Approximation

- Taking a second-order approximation, e.g. schmitt and grohe 2004, gives a system in the form
- $$x_t = d + Ex_{t-1} + Fu_t + \left(I \otimes x'_{t-1}\right) Gx_{t-1} + \left(I \otimes x'_{t-1}\right) Hu_t + \left(I \otimes u'_t\right) Ju_t.$$

  E and F are the coefficient matrices from the first-order approximation of the model,
- d is a correction for volatility that does not appear in the first-order approximation
- The matrix G contains coefficients on the squares and cross-products of the model's endogenous variables;
- the matrix H has coefficients for interaction terms between lagged endogenous variables and current-period shocks;
- and the matrix J relates to squares and cross-products of the shocks.

Auxiliary Disturbance Particle Filter

We propose a new particle filter for this problem. The auxiliary disturbance particle. See Hall, Pitt and Kohn (2013).

Application: Asset pricing with Habits

- the representative agent's consumption process is

$$\Delta \log C_t = g + \nu_t \ , \tag{5}$$

$\nu \sim N(0, \sigma^2)$. g is the long run average growth rate of real consumption, and ν_t is a transitory shock to income in period t.

Application: Asset pricing with Habits

- the representative agent's consumption process is

$$\Delta \log C_t = g + \nu_t \ , \tag{5}$$

$\nu \sim N(0, \sigma^2)$. g is the long run average growth rate of real consumption, and ν_t is a transitory shock to income in period t.

- The agent's utility function is given by

$$U_t = E_t \sum_{h=0}^{\infty} (\beta_{t+h})^h \frac{(C_{t+h} - X_{t+h})^{1-\gamma}}{1-\gamma} \ , \tag{6}$$

where X_t is the (external) habit stock, interpreted as the minimum level of consumption required to maintain a well-defined utility (i.e., the household must ensure that $C_t > X_t$).

Application: Asset pricing with Habits

- the representative agent's consumption process is

$$\Delta \log C_t = g + \nu_t \ , \tag{5}$$

  $\nu \sim N(0, \sigma^2)$. g is the long run average growth rate of real consumption, and ν_t is a transitory shock to income in period t.
- The agent's utility function is given by

$$U_t = E_t \sum_{h=0}^{\infty} (\beta_{t+h})^h \frac{(C_{t+h} - X_{t+h})^{1-\gamma}}{1-\gamma} \ , \tag{6}$$

  where X_t is the (external) habit stock, interpreted as the minimum level of consumption required to maintain a well-defined utility (i.e., the household must ensure that $C_t > X_t$).
- The intertemporal discount factor β_t measures impatience to consume in period t, and the parameter γ controls risk aversion.

Application: Asset pricing with Habits II

- To close the model, one must specify a law of motion for X_t. Convenient to do so by defining the surplus consumption ratio S_t, and the deviation $\widetilde{s}_t$ of $\log S_t$ from its mean $\overline{S}$, by

$$S_t = (C_t - X_t)/C_t \quad \text{and} \quad \widetilde{s}_t = \log S_t - \log \overline{S}$$

Application: Asset pricing with Habits II

- To close the model, one must specify a law of motion for X_t. Convenient to do so by defining the surplus consumption ratio S_t, and the deviation $\widetilde{s}_t$ of $\log S_t$ from its mean $\overline{S}$, by

$$S_t = (C_t - X_t)/C_t \quad \text{and} \quad \widetilde{s}_t = \log S_t - \log \overline{S}$$

- The law of motion of $\widetilde{s}_t$ is assumed to be

$$\widetilde{s}_t = \phi \widetilde{s}_{t-1} + (\overline{S}^{-1} \sqrt{1 - 2\widetilde{s}_{t-1}} - 1)\nu_t \ , \tag{7}$$

Application: Asset pricing with Habits II

- To close the model, one must specify a law of motion for X_t. Convenient to do so by defining the surplus consumption ratio S_t, and the deviation $\widetilde{s}_t$ of $\log S_t$ from its mean $\overline{S}$, by

$$S_t = (C_t - X_t)/C_t \quad \text{and} \quad \widetilde{s}_t = \log S_t - \log \overline{S}$$

- The law of motion of $\widetilde{s}_t$ is assumed to be

$$\widetilde{s}_t = \phi \widetilde{s}_{t-1} + (\overline{S}^{-1}\sqrt{1 - 2\widetilde{s}_{t-1}} - 1)\nu_t \ , \tag{7}$$

- Then the equilibrium price-dividend ratio of a financial asset satisfies

$$\frac{P_t}{D_t} = \beta_t \mathbb{E}_t \left[\exp\left[\gamma(\widetilde{s}_t - \widetilde{s}_{t+1}) + (1-\gamma)(g + \nu_{t+1})\right] \left(1 + \frac{P_{t+1}}{D_{t+1}}\right) \right] , \tag{8}$$

where β_t is the intertemporal discount factor in period t.

Data

- We apply the model to observations of growth in the S&P500 price-dividend ratio and US consumption using quarterly observations from 1950 to 2011, a total of 248 datapoints.

Data

- We apply the model to observations of growth in the S&P500 price-dividend ratio and US consumption using quarterly observations from 1950 to 2011, a total of 248 datapoints.
- The S&P500 series is from shiller 2006, while the consumption series is the seasonally adjusted real personal consumption expenditure series from the Bureau of Economic Analysis (series code PCECC96).