A Flexible State-Space Model with Application to Stochastic Volatility

Christian Gourieroux and Yang Lu

University of Toronto/CREST, and Aix-Marseille University

January, 2017

INTRODUCTION

State-space models

- Several terminologies: latent factor, dynamic frailty, dynamic random effect/unobserved heterogeneity, etc
- An alternative to observation-driven models (e.g. ARMA, ARCH).
- Several advantages:
  - More intuitive.
  - Suitable for irregular observations/missing data.
  - Simple stationary condition.
  - (In finance) suitable for derivative pricing.

Potential applications: time series/panel data

There is an increasing need for flexible models:

- (stochastic probability) event data, such as credit defaults [Duffie et al. (2009)].
- (stochastic intensity) count data, e.g. auto-mobile insurance; count of transactions.
- (stochastic volatility) income dynamics [Jensen and Shore (2011)].
- dynamic discrete choice data [Kasahara and Shimotsu (2009); Hu and Shum (2012): state variable $=$ taste, or belief]

However, state-space models have a high computational cost.

We propose a large family of state-space models that is:

- Tractable for estimation/prediction/filtering/smoothing.
- Flexible.

The computational gain: a foretaste

- Standard methods involve simulation: e.g. particle filter/MCMC/indirect inference. They are:
  - Slow
  - The likelihood function is computed with sampling error.
- Our model is associated with *simulation-free* composite likelihood estimation (CL), as well as simple procedures for forecasting and filtering.
- This is due to an endogenous switching regime interpretation of the model.

The stochastic volatility (SV) literature

Our model contributes to the SV literature by capturing:

- Heavy-tailed return.
- Conditional skewness, i.e. volatility feedback.
- Time ir-reversibility.

THE MODEL

The variables

- X_t: the univariate, state variable (latent factor) with domain $\mathcal{X}$ such as $\mathbb{R}$, $\mathbb{R}_{>0}$, or a bounded interval such as $]0, 1[$,
- Y_t the observable variable with domain $\mathcal{Y}$ such as $\mathbb{R}$, $\mathbb{R}_{>0}$, $\mathbb{N}$, or $\{0, 1, ..., N\}$.

The state-space representation

$$
\begin{aligned}
X_t | \underline{Y_{t-1}}, \underline{X_{t-1}} &\sim l(x_t | \underline{x_{t-1}}), \\
Y_t | \underline{Y_{t-1}}, \underline{X_t} &\sim l(y_t | x_t, \underline{y_{t-1}}),
\end{aligned}
$$

In terms of the causal scheme:

$$
\begin{array}{ccccccccc}
\cdots & \longrightarrow & X_{t-1} & \longrightarrow & X_t & \longrightarrow & X_{t+1} & \longrightarrow & \cdots \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
\cdots & \longrightarrow & Y_{t-1} & \longrightarrow & Y_t & \longrightarrow & Y_{t+1} & \longrightarrow & \cdots
\end{array}
$$

The dynamics of the state variable

We assume that process (X_t) is Markov and stationary. Its dynamics can be specified by:

- either: conditional transition density $l(x_{t+1}|x_t)$.
- or: joint density $l(x_t, x_{t+1})$.

We use the second approach.

The joint density

Assume:

$$f(x_t, x_{t+1}) = \frac{1}{M}\phi(x_t)\phi(x_{t+1})\Big[\sum_{j=0}^{J}\sum_{k=0}^{J} b_{j,k} x_t^j x_{t+1}^k\Big]^2,$$

where $\phi(\cdot)$ is a benchmark density; the coefficients $B = (b_{j,k})$ are real.

The normalization constant M is equal to:

$$M = \sum_{j_1,k_1,j_2,k_2=0}^{J} b_{j_1,k_1} b_{j_2,k_2} \mu_{j_1+j_2} \mu_{k_1+k_2}$$

where $\mu_k := \int x^k \phi(x) \mathrm{d}x$, $k = 0, ..., 2J$.

The joint density

Assume:

$$f(x_t, x_{t+1}) = \frac{1}{M}\phi(x_t)\phi(x_{t+1})\Big[\sum_{j=0}^{J}\sum_{k=0}^{J} b_{j,k} x_t^j x_{t+1}^k\Big]^2,$$

where $\phi(\cdot)$ is a benchmark density; the coefficients $B = (b_{j,k})$ are real.

The normalization constant M is equal to:

$$M = \sum_{j_1,k_1,j_2,k_2=0}^{J} b_{j_1,k_1} b_{j_2,k_2} \mu_{j_1+j_2} \mu_{k_1+k_2}$$

where $\mu_k := \int x^k \phi(x)\mathrm{d}x$, $k = 0, ..., 2J$.

A compact matrix form

We introduce matrix D by

$$d_{j,k} = \mu_j \mu_k \sum_{\substack{j_1+j_2=j \\ 0 \le j_1, j_2 \le J}} \sum_{\substack{k_1+k_2=k \\ 0 \le k_1, k_2 \le J}} b_{j_1,k_1} b_{j_2,k_2}, \qquad 0 \le j, k \le 2J$$

Then we can rewrite $M = e'De$, with $e = (1, 1, ..., 1)'$ the unitary vector. Similarly,

$$f(x_t, x_{t+1}) = \phi(x_t)\phi(x_{t+1}) \frac{U'(x_t) D U(x_{t+1})}{e'De},$$

where vector function U is defined by:

$$U_j(x) = \frac{x^j}{\mu_j}, \qquad \forall j = 0, 1, ..., 2J,$$

that satisfies: $\int \phi(x) U(x) \mathrm{d}x = e$.

A compact matrix form

We introduce matrix D by

$$d_{j,k} = \mu_j \mu_k \sum_{\substack{j_1+j_2=j \\ 0 \le j_1, j_2 \le J}} \sum_{\substack{k_1+k_2=k \\ 0 \le k_1, k_2 \le J}} b_{j_1,k_1} b_{j_2,k_2}, \qquad 0 \le j, k \le 2J$$

Then we can rewrite $M = e'De$, with $e = (1, 1, ..., 1)'$ the unitary vector. Similarly,

$$f(x_t, x_{t+1}) = \phi(x_t)\phi(x_{t+1}) \frac{U'(x_t)DU(x_{t+1})}{e'De},$$

where vector function U is defined by:

$$U_j(x) = \frac{x^j}{\mu_j}, \qquad \forall j = 0, 1, ..., 2J,$$

that satisfies: $\int \phi(x)U(x)\mathrm{d}x = e$.

The stationarity condition

Entries of D should satisfy the condition that X_t and X_{t+1} have equal margins. Since

$$
\begin{aligned}
l(x_t) &= \int f(x_t, x_{t+1})\mathrm{d}x_{t+1} = \phi(x_t)\frac{U'(x_t)De}{e'De}, \\
l(x_{t+1}) &= \int f(x_t, x_{t+1})\mathrm{d}x_t = \phi(x_{t+1})\frac{e'DU(x_{t+1})}{e'De},
\end{aligned}
$$

equalising the two densities leads to:

$$(D - D')e = 0.$$

Conditional density

From the joint distribution

$$f(x_t, x_{t+1}) = \phi(x_t)\phi(x_{t+1})\frac{U'(x_t)DU(x_{t+1})}{e'De}$$

and the marginal

$$l(x_t) = \phi(x_t)\frac{U'(x_t)De}{e'De},$$

we get the conditional distribution :

$$l(x_{t+1}|x_t) = \frac{f(x_t, x_{t+1})}{l(x_t)} = \phi(x_{t+1})\frac{U'(x_t)DU(x_{t+1})}{U'(x_t)De}.$$

This formula is easily generalized to higher horizons.

Conditional density at higher horizon

Proposition 1

The conditional distribution of X_{t+h} given X_t is:

$$l(x_{t+h}|x_t) = \phi(x_{t+h})U'(x_t)\frac{D\Pi^{h-1}U(x_{t+h})}{U'(x_t)De},$$

where the $(2J+1) \times (2J+1)$ matrix Π is defined by:

$$\Pi = \int \phi(x)\frac{U(x)U'(x)D}{U'(x)De}\mathrm{d}x.$$

The eigenvectors of Π

Proposition 2

- *(Right eigenvector) The components of each row of Π sum up to one: $\Pi e = e$.*
- *(Left eigenvector) $(De)'\Pi = (De)'$.*

Thus Π has a transition matrix (of a Markov chain) interpretation and De has a stationary distribution interpretation.

The ergodicity condition

Proposition 3

The state process (X_t) is weakly ergodic if 1 is a simple eigenvalue of matrix Π, and all other eigenvalues are smaller than 1 in modulus.

THE SWITCHING REGIME REPRESENTATION

Remind: $I(x_{t+1}|x_t) = \left[\frac{U(x_t)'D}{U'(x_t)De}\right]\left[\phi(x_{t+1})U(x_{t+1})\right]$.

Proposition 4

If (X_t) is positive, and D nonnegative, then process (X_t) admits a switching regime (S_t), with $2J+1$ possible values:

$$\left(\mathbb{P}[S_t = 0|\underline{S_{t-1}}, \underline{X_t}], ..., \mathbb{P}[S_t = 2J|\underline{S_{t-1}}, \underline{X_t}]\right)' = \frac{U'(X_t)D}{U'(X_t)De},$$

and

$$I(x_{t+1}|S_t = i, \underline{x_t}) \sim \phi(x_{t+1})U_{s_t}(x_{t+1}), \quad ß = 0, ..., 2J.$$

To summarize, we have the following causal scheme:

$$\ldots X_{t-1} \to S_{t-1} \to X_t \to S_t \to X_{t+1} \to S_{t+1} \ldots.$$

Remind: $l(x_{t+1}|x_t) = \left[\frac{U(x_t)'D}{U'(x_t)De}\right]\left[\phi(x_{t+1})U(x_{t+1})\right]$.

Proposition 4

If (X_t) is positive, and D nonnegative, then process (X_t) admits a switching regime (S_t), with $2J+1$ possible values:

$$\left(\mathbb{P}[S_t = 0|\underline{S_{t-1}}, \underline{X_t}], ..., \mathbb{P}[S_t = 2J|\underline{S_{t-1}}, \underline{X_t}]\right)' = \frac{U'(X_t)D}{U'(X_t)De},$$

and

$$l(x_{t+1}|S_t = i, \underline{x_t}) \sim \phi(x_{t+1})U_{s_t}(x_{t+1}), \quad ß = 0, ..., 2J.$$

To summarize, we have the following causal scheme:

$$\ldots X_{t-1} \to S_{t-1} \to X_t \to S_t \to X_{t+1} \to S_{t+1} \ldots.$$

Interpretation of Π

Corollary 1

The embedded latent process (S_t) *is also a Markov chain with transition matrix* Π*:*

$$\Pi_{i,j} = \mathbb{P}[S_{t+1} = j | S_t = i], \quad \forall i, j = 0, ...2J.$$

Comparison with standard Markov switching models

This causality structure is different from the standard Markov switching model:

$$\begin{array}{ccccccc} \dots & S_{t-1} & \longrightarrow & S_t & \longrightarrow & S_{t+1} & \dots \\ & \downarrow & & \downarrow & & \downarrow & \\ \dots & X_{t-1} & & X_t & & X_{t+1} & \dots \end{array}$$

Indeed,

- our switching regime S_t may not exist (D can have negative entries).
- When it exists, its transition probabilities are endogenous: $l(S_{t+1}|S_t, X_t)$ depends on X_t.
- However, the two models share the common point that the estimation/prediction/filtering is simple.

Comparison with standard Markov switching models

This causality structure is different from the standard Markov switching model:

$$\begin{array}{ccccccc} \dots & S_{t-1} & \longrightarrow & S_t & \longrightarrow & S_{t+1} & \dots \\ & \downarrow & & \downarrow & & \downarrow & \\ \dots & X_{t-1} & & X_t & & X_{t+1} & \dots \end{array}$$

Indeed,

- our switching regime S_t may not exist (D can have negative entries).
- When it exists, its transition probabilities are endogenous: $l(S_{t+1}|S_t, X_t)$ depends on X_t.
- However, the two models share the common point that the estimation/prediction/filtering is simple.

Similar processes: ARG

$$\ldots X_{t-1} \to S_{t-1} \to X_t \to S_t \to X_{t+1} \to S_{t+1} \ldots .$$

The Autoregressive Gamma (ARG) process is the exact time discretization of the Cox-Ingersoll-Ross process:

$$\mathrm{d}X_t = a(b - X_t)\mathrm{d}t + \sigma\sqrt{X_t}\mathrm{d}W_t.$$

It can be shown that $(X_t)_{t\in\mathbb{N}}$ has the representation:

- Given X_t, latent regime S_t is Poisson distributed $\mathcal{P}(\rho X_t)$,
- Given regime S_t, variable X_{t+1} is gamma distributed $\gamma(\delta + S_t, c)$,

where $\rho = e^{-a}, \delta = \frac{2ab}{\sigma^2}, c = \frac{1-e^{-a}}{2a}\sigma^2$.

TIME REVERSIBILITY

- A process (X_t) is (time) reversible if (X_t) and (X_{-t}) have the same dynamics.
- In the Markov case, time reversibility is equivalent to $f(x_t, x_{t+1}) = f(x_{t+1}, x_t)$.
- Many time series models, such as univariate discretized diffusions (CIR/ARG), normal ARMA models, are reversible.
- Most existing SV models are reversible.
- However, theory predicts ir-reversibility [Maskin and Tirole (1988)].
- Linked to the bubble phenomenon.

Existing empirical studies

- The reversibility is also rejected by empirical tests for many economic and financial data [see e.g. Ramsey and Rothman (1996); Chen et al. (2000)].
- However, once tested, no appropriate models are provided to account for ir-reversibility.
- Our model is capable of capturing time ir-reversibility.
- First, let us look at the condition of reversibility.

Condition of time reversibility

Proposition 5

Process (X_t) is reversible if and only if D is symmetric, or equivalently, B is symmetric or antisymmetric.

If D is symmetric, condition $(D - D')e = 0$ is automatic.

Can our model accommodate for non symmetric D?

Yes so long as $(D - D')e = 0$ is satisfied.

Condition of time reversibility

Proposition 5

Process (X_t) is reversible if and only if D is symmetric, or equivalently, B is symmetric or antisymmetric.

If D is symmetric, condition $(D - D')e = 0$ is automatic.

Can our model accommodate for non symmetric D?

Yes so long as $(D - D')e = 0$ is satisfied.

Symmetric-Antisymmetric decomposition

Each square matrix B can be decomposed into $B = B_1 + B_2$, with $B_1 = \frac{1}{2}(B + B')$ symmetric and $B_2 = \frac{1}{2}(B - B')$ antisymmetric.

Let us derive the corresponding decomposition of D, defined by:

$$d_{j,k} = \mu_j \mu_k \sum_{\substack{j_1+j_2=j \\ 0 \le j_1, j_2 \le J}} \sum_{\substack{k_1+k_2=k \\ 0 \le k_1, k_2 \le J}} b_{j_1,k_1} b_{j_2,k_2}.$$

Decomposing D

$$
\begin{aligned}
d_{j,k} &= \mu_j \mu_k \sum_{j_1,j_2} \sum_{k_1,k_2} (b_{1,j_1,k_1} + b_{2,j_1,k_1})(b_{1,j_2,k_2} + b_{2,j_2,k_2}) \\
&= \underbrace{\mu_j \mu_k \sum_{j_1,j_2} \sum_{k_1,k_2} (b_{1,j_1,k_1} b_{1,j_2,k_2} + b_{2,j_1,k_1} b_{2,j_2,k_2})}_{=D_1\text{, symmetric part}} \\
&\quad + \underbrace{2\mu_j \mu_k \sum_{j_1,j_2} \sum_{k_1,k_2} b_{1,j_1,k_1} b_{2,j_2,k_2},}_{=D_2\text{, antisymmetric part}} \\
&= D_1 + D_2
\end{aligned}
$$

Moreover, $(D - D')e = 2D_2 e = 0$. These are **linear** constraints on B_2, where B_1 is free parameter.

Example: $J = 1$

$$f(x_t, x_{t+1}) = \frac{1}{M}\phi(x_t)\phi(x_{t+1})\Big(b_{00} + b_{10}x_t + b_{01}x_{t+1} + b_{11}x_t x_{t+1}\Big)^2.$$

$$D = \begin{bmatrix} b_{00}^2 & 2b_{00}b_{01}\mu_1 & b_{01}^2\mu_2 \\ 2b_{00}b_{10}\mu_1 & 2b_{00}b_{11}\mu_1^2 + 2b_{01}b_{10}\mu_1^2 & 2b_{01}b_{11}\mu_1\mu_2 \\ b_{10}^2\mu_2 & 2b_{10}b_{11}\mu_1\mu_2 & b_{11}^2\mu_2^2 \end{bmatrix}.$$

The equal margin condition $(D - D')e = 0$ becomes:

$$\begin{aligned} (b_{01} - b_{10})(2b_{00}\mu_1 + \mu_2(b_{10} + b_{01})) &= 0, \\ (b_{01} - b_{10})\mu_2(b_{01} + b_{10} + 2b_{11}\mu_1) &= 0, \\ (b_{01} - b_{10})\mu_1(b_{00} - b_{11}\mu_2) &= 0. \end{aligned}$$

The non-parametric background

Any joint density $f(x_t, x_{t+1})$ can be decomposed with respect to any benchmark product density $\phi(x_t)\phi(x_{t+1})$ into:

$$\sqrt{\frac{f(x_t, x_{t+1})}{\phi(x_t)\phi(x_{t+1})}} = \sum_{i,j=0}^{\infty} a_{i,j} P_i(x_t) P_j(x_{t+1}),$$

where (P_j) are orthonormal polynomials for measure $\phi(x)\mathrm{d}x$:

$$\int \phi(x) P_i(x) P_j(x) \mathrm{d}x = \mathbb{1}_{i=j}.$$

Then the expansion is truncated at order J, and renormalised:

$$f_J = \frac{1}{M_J} \phi(x_t)\phi(x_{t+1}) \Big[\sum_{i,j=0}^{J} a_{i,j} P_i(x_t) P_j(x_{t+1}) \Big]^2.$$

The distributional flexibility

The density f_J is flexible in the sense that:

Proposition 6

For any benchmark density ϕ, the sequence of densities f_J approximates f arbitrarily well in terms of the Hellinger distance, when J goes to infinity:

$$\iint \left| \sqrt{f_J(x_t, x_{t+1})} - \sqrt{f(x_t, x_{t+1})} \right|^2 \mathrm{d}x_t \mathrm{d}x_{t+1} \longrightarrow 0.$$

Comparison with the literature

- The literature [Jarrow and Rudd (1982); Aït-Sahalia (2002); Filipović et al. (2013)] has also proposed:

$$\frac{f(x_t, x_{t+1})}{\phi(x_t)\phi(x_{t+1})} = \sum_{i,j=0}^{\infty} c_{i,j} P_i(x_t) P_j(x_{t+1}),$$

- when ϕ is normal, we get the Edgeworth expansion, or Gram-Schmidt (Hermit) expansion.
- Some downsides:
  - the truncated density of these authors is NOT positive (arbitrary opportunities, impossibility of evaluating the quality of approximation).
  - requires integrability conditions that are often violated [see e.g. Aït-Sahalia (2002)]

Comparison with the literature

- The literature [Jarrow and Rudd (1982); Aït-Sahalia (2002); Filipović et al. (2013)] has also proposed:

$$\frac{f(x_t, x_{t+1})}{\phi(x_t)\phi(x_{t+1})} = \sum_{i,j=0}^{\infty} c_{i,j} P_i(x_t) P_j(x_{t+1}),$$

- when ϕ is normal, we get the Edgeworth expansion, or Gram-Schmidt (Hermit) expansion.
- Some downsides:
  - the truncated density of these authors is NOT positive (arbitrary opportunities, impossibility of evaluating the quality of approximation).
  - requires integrability conditions that are often violated [see e.g. Aït-Sahalia (2002)]

FORECASTING AND ESTIMATION

Proposition 7

$$l(y_t|\underline{y_{t-1}}) = P'(\underline{y_{t-1}})g(y_t|\underline{y_{t-1}})$$

where

$$g(y_t|\underline{y_{t-1}}) = \Big[\int_0^{\infty} l(y_t|x_t, \underline{y_{t-1}})\phi(x_t)U(x_t)\mathrm{d}x_t\Big].$$

In terms of latent regime:

- $P'(\underline{y_{t-1}})$ is the vector of probabilities of S_t given past observations $\underline{y_{t-1}}$.
- $g(y_t|\underline{y_{t-1}})$ is the vector of conditional densities of $y_t|\underline{y_{t-1}}, s_t$.

It remains to compute recursively $P'(\underline{y_{t-1}})$.

Proposition 8

The vector $P'(\underline{y_{t-1}})$ is computed recursively by:

$$P'(\underline{y_t}) = P'(\underline{y_{t-1}})\Pi(\underline{y_t}),$$

$$\textit{with initial condition } P'(\underline{y_0}) = \frac{e'D}{e'De},$$

with matrix $\Pi(\underline{y_t})$ given by:

$$\Pi(\underline{y_t}) := \frac{1}{l(y_t|\underline{y_{t-1}})} \int \phi(x_t) \frac{U(x_t)U'(x_t)D}{U'(x_t)De} l(y_t|x_t, \underline{y_{t-1}}) \mathrm{d}x_t.$$

Interpretation

- $\Pi(\underline{y_t})$ is the “endogenous transition matrix" of the chain S_t given $\underline{y_t}$:
$$\pi_{i,j}(\underline{y_t}) = \mathbb{P}[S_{t+1} = j | S_t = i, \underline{y_t}]$$
- Similarly, we have closed form formulas for the filtering density $l(x_t|\underline{y_t})$, and smoothing density $l(x_t|\underline{y_T})$.

The model without feedback

The predictive density $l(y_t|\underline{y_{t-1}})$ can also be used to compute the likelihood function.
Nevertheless, in a special case, the estimation can be further simplified. Let us assume:

$$l(y_t|\underline{y_{t-1}}, x_t) = l(y_t|x_t).$$

That is:

$$\begin{array}{ccccccc}
\cdots & X_{t-1} & \longrightarrow & X_t & \longrightarrow & X_{t+1} & \cdots \\
 & \downarrow & & \downarrow & & \downarrow & \\
\cdots & Y_{t-1} & & Y_t & & Y_{t+1} & \cdots
\end{array}$$

Proposition 9

Under this assumption, we have, for each $h \geq 1$:

$$l(y_t, y_{t+h}) = g'(y_t)\frac{D\Pi^{h-1}}{e'De}g(y_{t+h}),$$

where

$$g(y_t) = \int l(y_t|x_t)\phi(x_t)U(x_t)\mathrm{d}x_t.$$

The (pairwise) composite likelihood (CL) estimation

The maximum composite likelihood estimator is:

$$\hat{\theta} = \arg\max_{\theta} \ell_{CL}(\theta)$$

where

$$\ell_{CL}(\theta) = \sum_{t=1}^{T} \sum_{h=1}^{\min(m,T-t)} \omega_h \log l(y_t, y_{t+h}|\theta),$$

where weight ω_h is nonnegative.
Thus the pairwise CL function evaluates the joint densities of all neighbouring pairs (Y_t, Y_{t+h}) with distance non larger than m.

The advantages of the CL

- Extremely low computational cost (simulation-free), when $g(y_t)$ has closed form.
- The latter is rendered possible by appropriate forms for ϕ and $l(y_t|x_t)$.
- Consistent, asymptotically normal and quite efficient [see Varin et al. (2011)].

A STOCHASTIC VOLATILITY APPLICATION

Model 1

Assume daily return Y_t satisfies

$$y_t = \frac{1}{\sqrt{x_t}} \epsilon_t,$$

where ϵ_t is i.i.d. standard normal.

We use gamma benchmark density:

$$\phi(x_t) = \frac{1}{\Gamma(\alpha) c^\alpha} x_t^{\alpha-1} e^{-x_t/c}.$$

Then $\phi(x_t)U(x_t)$ are gamma densities, and the components of $g(y_t) = \int l(y_t|x_t)\phi(x_t)U(x_t)\mathrm{d}x_t$ are:

$$g_j(y_t) = \frac{\Gamma(\alpha + j + \frac{1}{2})}{\Gamma(\alpha + j)\sqrt{2\pi} c^{\alpha+j} (\frac{y_t^2}{2} + \frac{1}{c})^{\alpha+j+\frac{1}{2}}}, \qquad \forall j = 0, ..., 2J.$$

In other words these are re-scaled student densities. That is, Y_t is conditionally heavy tailed.

Model 1

Assume daily return Y_t satisfies

$$y_t = \frac{1}{\sqrt{x_t}}\epsilon_t,$$

where ϵ_t is i.i.d. standard normal.

We use gamma benchmark density:

$$\phi(x_t) = \frac{1}{\Gamma(\alpha)c^\alpha} x_t^{\alpha-1} e^{-x_t/c}.$$

Then $\phi(x_t)U(x_t)$ are gamma densities, and the components of $g(y_t) = \int l(y_t|x_t)\phi(x_t)U(x_t)\mathrm{d}x_t$ are:

$$g_j(y_t) = \frac{\Gamma(\alpha + j + \frac{1}{2})}{\Gamma(\alpha + j)\sqrt{2\pi}c^{\alpha+j}(\frac{y_t^2}{2} + \frac{1}{c})^{\alpha+j+\frac{1}{2}}}, \qquad \forall j = 0, ..., 2J.$$

In other words these are re-scaled student densities. That is, Y_t is conditionally heavy tailed.

Model 2: Incorporating conditional skewness

The normal distribution of the error ϵ_t is symmetric. Let us replace its density by:

$$h(\epsilon) = \frac{1}{M_\epsilon}\psi(\epsilon)(1 + \sum_{i=1}^{I}\beta_i\epsilon^i)^2,$$

where $\psi(\epsilon) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\epsilon^2}{2}}$.

- Same spirit as the joint density of (X_t, X_{t+1}).
- Thus y_t is skewed: a flexible alternative to the skewed student distribution [Fernández and Steel (1998)].
- Thus we have *volatility feedback*, i.e. negative/positive past returns have different impacts on future returns.

Other models

As a benchmark, we also estimate Model M3, where (X_t) is autoregressive gamma (ARG).

Finally, Model M1, M2, M3 assume symmetric B, thus reversible dynamics for (X_t). Let us consider Model M4, with non symmetric B.

The data

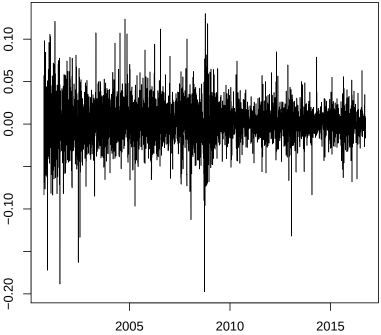
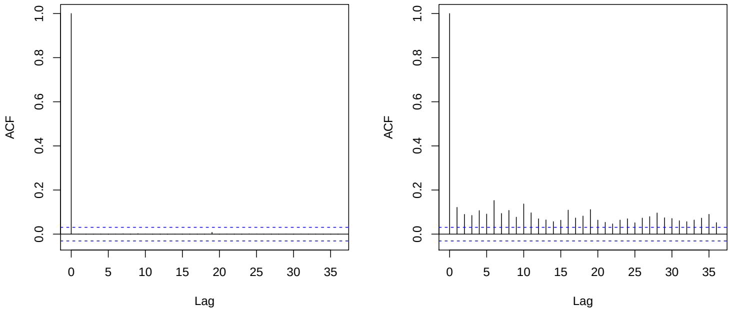


Figure: Daily return of the Apple stock between 2000/10/2 and 2016/9/29.

When the error (ϵ_t) has unitary variance, $\mathbb{E}[Y_t^2|X_t] = \frac{1}{X_t}$, thus $1/Y_t^2$ is a proxy of the volatility.



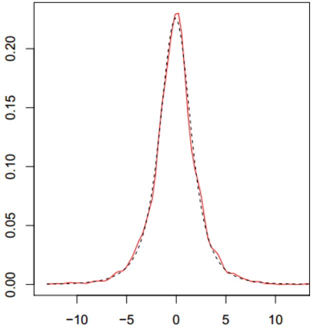
FigureHistorical autocorrelation function of $1/Y_t^2$ (left panel) and of Y_t^2, on the right panel

Comparison of the models

Model	M1 $J = 2$	M2 $J = 3$	M3 $--$	M4 $J = 4$	Real data
Error	Gaussian	skewed $I = 1$	skewed $I = 1$	skewed $I = 2$	$--$ $--$
Reversible	yes	yes	no	no	$--$
$\mathbb{E}[y_t]$	0	0.96	0.10	0.11	0.11
$\mathbb{E}[y_t^2]$	5.0	5.6	6.1	6.1	6.0
$\mathbb{E}[y_t^3]$	0	-0.95	-0.96	-1.21	-1.40
$\mathbb{E}[y_t^4]$	205	298	303	314	304
$\text{corr}[y_t^2, y_{t+1}^2]$	0.043	0.054	0.0001	0.09	0.12
ℓ_{CL}	-74356	-74105	-73861	-73407	$--$
AIC_{CL}	148726	148232	147756	146854	$--$

- Generically, increasing J or I both improves the fit, in terms of the “composite AIC".
- The model with ARG latent factor cannot capture the persistence of the volatility.

Model implied marginal density Vs kernel density estimate



The time ir-reversibility

For model M4,

- The antisymmetric matrix D_2 is non zero, which confirms the time non reversibility of the Apple return data.
- Nevertheless, the largest entry of $|D_1|$ is 1.47, whereas the largest (in absolute value) entry of $|D_2|$ is 0.122, which is significantly smaller than 1.47.

CONCLUSION

The contribution of the paper was twofold:

- To the state-space literature: a class of models based on a new specification for (univariate) state process, with
    - Distributional flexibility
    - Endogenous switching regime interpretation
    - Simulation-free prediction/filtering/estimation procedures.
- To the SV literature :
    - Captures *i*) heavy tail, *ii*) volatility feedback, *iii*) time ir-reversible dynamics.
    - Empirical illustration suggests a much better fit.

THANKS FOR YOUR ATTENTION.

An orthogonal expansion

Assume that there exists an orthonormal basis of polynomials $P_i(x)$, with $deg(P_i) = i$ for the $L^2(\phi(x)\mathrm{d}x)$. Then $(P_i(x_t)P_j(x_{t+1}))_{i,j}$ is an orthonormal basis for the product space $L^2(\phi(x_t)\phi(x_{t+1})\mathrm{d}x_t\mathrm{d}x_{t+1})$.

Since $f(x_t, x_{t+1})$ integrates to unity:

$$\iint \Big(\sqrt{\frac{f(x_t, x_{t+1})}{\phi(x_t)\phi(x_{t+1})}}\Big)^2 \phi(x_t)\phi(x_{t+1})\mathrm{d}x_t\mathrm{d}x_{t+1} = 1,$$

$\sqrt{\frac{f(x_t,x_{t+1})}{\phi(x_t)\phi(x_{t+1})}}$ belongs to the $L^2(\phi(x_t)\phi(x_{t+1})\mathrm{d}x_t\mathrm{d}x_{t+1})$. Thus we get the following orthonormal decomposition:

$$\sqrt{\frac{f(x_t, x_{t+1})}{\phi(x_t)\phi(x_{t+1})}} = \sum_{i,j=0}^{\infty} a_{i,j} P_i(x_t)P_j(x_{t+1}), \tag{1}$$

An orthogonal expansion

Assume that there exists an orthonormal basis of polynomials $P_i(x)$, with $deg(P_i) = i$ for the $L^2(\phi(x)\mathrm{d}x)$. Then $(P_i(x_t)P_j(x_{t+1}))_{i,j}$ is an orthonormal basis for the product space $L^2(\phi(x_t)\phi(x_{t+1})\mathrm{d}x_t\mathrm{d}x_{t+1})$.

Since $f(x_t, x_{t+1})$ integrates to unity:

$$\iint \Big(\sqrt{\frac{f(x_t, x_{t+1})}{\phi(x_t)\phi(x_{t+1})}}\Big)^2 \phi(x_t)\phi(x_{t+1})\mathrm{d}x_t\mathrm{d}x_{t+1} = 1,$$

$\sqrt{\frac{f(x_t,x_{t+1})}{\phi(x_t)\phi(x_{t+1})}}$ belongs to the $L^2(\phi(x_t)\phi(x_{t+1})\mathrm{d}x_t\mathrm{d}x_{t+1})$. Thus we get the following orthonormal decomposition:

$$\sqrt{\frac{f(x_t, x_{t+1})}{\phi(x_t)\phi(x_{t+1})}} = \sum_{i,j=0}^{\infty} a_{i,j} P_i(x_t)P_j(x_{t+1}), \tag{1}$$

An orthogonal expansion

Assume that there exists an orthonormal basis of polynomials $P_i(x)$, with $deg(P_i) = i$ for the $L^2(\phi(x)\mathrm{d}x)$. Then $(P_i(x_t)P_j(x_{t+1}))_{i,j}$ is an orthonormal basis for the product space $L^2(\phi(x_t)\phi(x_{t+1})\mathrm{d}x_t\mathrm{d}x_{t+1})$.

Since $f(x_t, x_{t+1})$ integrates to unity:

$$\iint \left(\sqrt{\frac{f(x_t, x_{t+1})}{\phi(x_t)\phi(x_{t+1})}}\right)^2 \phi(x_t)\phi(x_{t+1})\mathrm{d}x_t\mathrm{d}x_{t+1} = 1,$$

$\sqrt{\frac{f(x_t,x_{t+1})}{\phi(x_t)\phi(x_{t+1})}}$ belongs to the $L^2(\phi(x_t)\phi(x_{t+1})\mathrm{d}x_t\mathrm{d}x_{t+1})$. Thus we get the following orthonormal decomposition:

$$\sqrt{\frac{f(x_t, x_{t+1})}{\phi(x_t)\phi(x_{t+1})}} = \sum_{i,j=0}^{\infty} a_{i,j} P_i(x_t) P_j(x_{t+1}), \qquad (1)$$

Truncating this infinite expansion yields

$$f(x_t, x_{t+1}) \approx \phi(x_t)\phi(x_{t+1})\Big[\sum_{i,j=0}^{\infty} a_{i,j}P_i(x_t)P_j(x_{t+1})\Big]^2$$

After normalization, we get an approximated density:

$$f_J(x_t, x_{t+1}) = \frac{1}{M_J}\phi(x_t)\phi(x_{t+1})\Big[\sum_{i,j=0}^{J} a_{i,j}P_i(x_t)P_j(x_{t+1})\Big]^2,$$

with $M_J = \sum_{i,j=0}^{J} a_{i,j}^2$.

Finally, since the polynomial P_i has degree i, the term $\sum_{i,j=0}^{J} a_{i,j}P_i(x_t)P_j(x_{t+1})$ can be alternative expressed as $\sum_{i,j=0}^{J} b_{i,j}x_t^i x_{t+1}^j$.

Comparison with Markov switching models

- Transition probability no longer endogenous
- Our model is also applicable to benchmark densities other than gamma.
- Identifiability problem: how to choose the component densities in a Markov switching model?
- No equivalent result on Hellinger distance for Markov switching models.

Filtering formula

Proposition 10

The filtering density of X_t given the observables $\underline{Y_t}$ is:

$$l(x_t|\underline{y_t}) = \phi(x_t)\frac{P'(\underline{y_{t-1}})U(x_t)l(y_t|x_t,\underline{y_{t-1}})}{l(y_t|\underline{y_{t-1}})}$$

Smoothing formula

Proposition 11

If we introduce vectors (Q_t) backwardly:

$$Q_T = g(y_T | \underline{y_{T-1}}),$$
$$Q_{t-1} = \Pi(\underline{y_{t-1}}) Q_t, \forall t < T,$$

then we have:

$$l(x_t | \underline{y_T}) = \frac{1}{P'(\underline{y_{t-1}}) g(y_t | \underline{y_{t-1}})} \frac{P'(\underline{y_{t-1}}) \left[\phi(x_t) \frac{U(x_t)U'(x_t)D}{U'(x_t)De} l(y_t | x_t) \right] Q_{t+1}}{P'(\underline{y_{t-1}}) \Pi(\underline{y_t}) Q_{t+1}}$$

Other conditions for ergodicity

Proposition 12

If the state process (X_t) is positive, then any of the two following conditions implies the ergodicity of (X_t):

- *all entries of D are nonnegative;*
- *D is symmetric.*

Finite dimensional dependence (FDD)

- In our model, the joint density is the sum of cross products of elementary functions of x_t and x_{t+1}.
- Such markov processes are of have *finite dimensional dependence* (FDD).
- FDD models have been shown to be a serious competitor [see e.g. Gouriéroux and Monfort (2015)] of affine term structure models [such as the CIR model].
- As affine models, our model allows for closed form pricing formulas, and offers distributional flexibility (essential to capture interest rates at a large number of horizons).

Aït-Sahalia, Y. (2002). Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach. *Econometrica*, 70(1):223–262.

Bauwens, L. and Hautsch, N. (2006). Stochastic Conditional Intensity Processes. *Journal of Financial Econometrics*, 4(3):450–493.

Chen, Y.-T., Chou, R. Y., and Kuan, C.-M. (2000). Testing Time Reversibility Without Moment Restrictions. *Journal of Econometrics*, 95(1):199–218.

Duffie, D., Eckner, A., Horel, G., and Saita, L. (2009). Frailty Correlated Default. *Journal of Finance*, 64(5):2089–2123.

Fernández, C. and Steel, M. F. (1998). On Bayesian Modeling of Fat Tails and Skewness. *Journal of the American Statistical Association*, 93(441):359–371.

Filipović, D., Mayerhofer, E., and Schneider, P. (2013). Density Approximations for Multivariate Affine Jump-Diffusion Processes. *Journal of Econometrics*, 176(2):93–111.

Gouriéroux, C. and Monfort, A. (2015). Pricing with Finite Dimensional Dependence. *Journal of Econometrics*, 187(2):408–417.

Hu, Y. and Shum, M. (2012). Nonparametric Identification of Dynamic Models with Unobserved State Variables. *Journal of Econometrics*, 171(1):32–44.

Jarrow, R. and Rudd, A. (1982). Approximate Option Valuation for Arbitrary Stochastic Processes. *Journal of Financial Economics*, 10(3):347–369.

Jensen, S. T. and Shore, S. H. (2011). Semiparametric Bayesian Modeling of Income Volatility Heterogeneity. *Journal of the American Statistical Association*, 106(496):1280–1290.

Kasahara, H. and Shimotsu, K. (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica*, 77(1):135–175.

Maskin, E. and Tirole, J. (1988). A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles. *Econometrica*, 56(3):571–599.

McCausland, W. J. (2007). Time Reversibility of Stationary Regular Finite-State Markov Chains. *Journal of Econometrics*, 136(1):303–318.

Ramsey, J. B. and Rothman, P. (1996). Time Irreversibility and Business Cycle Asymmetry. *Journal of Money, Credit and Banking*, 28(1):1–21.

Varin, C., Reid, N., and Firth, D. (2011). An Overview of Composite Likelihood Methods. *Statistica Sinica*, 21(1):5–42.