Massimo Guidolin
Massimo.Guidolin@unibocconi.it
Dept. of Finance

Università Commerciale
Luigi Bocconi

ADVANCED FINANCIAL ECONOMETRICS – PROF. MASSIMO GUIDOLIN©

LECTURE 4: AFFINE VS. NON-AFFINE DAPMS

OVERVIEW

1) Definition and spanning of the family
2) The jump diffusion continuous time case
3) Discrete time affine models
4) Transformation of affine models
5) GMM and QML estimation
6) MLE: Full-information, simulated, approximate
7) Characteristic function-based estimation
8) Hints to non-affine models (GARCH, SV, jumps, mixtures, etc.)

The affine family: Definition

- Are you looking for generally realistic, potentially complex and yet tractable time series models of financial returns? Then you are after affine models
  - This tractability derives from the knowledge of closed-form solutions to several "transforms" of affine processes which are Markov processes
  - Intuitively, an affine process Y is one for which the conditional mean and variance are affine functions of Y, but we shall see that the family is much wider and more interesting
  - They can appear as continuous time processes (possibly then discretized) or directly as discrete time processes
- Unless otherwise noted, we assume that Y is observable
- **Definition 7 [Markov process]** A process Y is Markov if, for any measurable function g: D → $\mathbb{R}$ and for any fixed times

THE AFFINE FAMILY: DEFINITION

t and s > t , $E_t[g(Y_s)] = h(Y_t)$ for some fnct *h*: D → $\mathbb{R}$

– In short, the conditional distribution of Y_s only depends on the current state Y_t
  - When the conditional distribution of Y depends on additional lags of Y , one can often expand the dimension of the state vector to obtain a new, first-order Markov process Y*
  - Even when, in continuous time, the conditional distribution of Y_s depends on a continuum of lagged values of Y_t, it is often possible to define a new state variable that captures this dependence

– Let's define CCFs and CMGFs: the conditional characteristic function (CCF) of a Markov process Y_T, conditioned on current and lagged information is given by the Fourier transform of its conditional density function (τ ≡ T – t)

$$\text{CCF}_t(\tau, u) \equiv E\left(e^{iu'Y_T} \mid Y_t\right), \quad u \in \mathbb{R}^N$$

$$= \int_{\mathbb{R}^N} f_Y(Y_T|Y_t;\gamma)e^{iu'Y_T}\, dY_T$$

$i = \sqrt{-1}$

THE AFFINE FAMILY: DEFINITION

– Similarly, the conditional moment-generating function (CMGF) is given by the Laplace transform of Y :

$$\text{CMGF}_t(\tau, u) \equiv E\left(e^{u'Y_T} \mid Y_t\right), \quad u \in \mathbb{R}^N$$

$$= \int_{\mathbb{R}^N} f_Y(Y_T|Y_t; \gamma) e^{u'Y_T} \, dY_T$$

- **Definition 8 [Affine process]** A Markov process Y is said to be an affine process if either its CCF or CMGF has the exponential affine form $\text{CCF}_t \text{ or } \text{CMGF}_t = e^{\phi_{0t} + \phi'_{Yt} Y_t}$

  where ϕ_{0t} and $\boldsymbol{\phi}_{Yt}$ are complex (real) coefficients indexed by t to allow for time dependence of moments

  – In continuous time, affine processes relate to special restrictions to be imposed on jump-diffusion processes

The affine family: Jump diffusion case

– A jump-diffusion process is a Markov process solving the stochastic differential equation

Standard N-dimensional Brownian motion

$$dY_t = \mu(Y_t, \gamma_0)\, dt + \sigma(Y_t, \gamma_0)\, dW_t + dZ_t$$

– Z_t is a pure-jump process whose jump amplitudes have a fixed probability distribution ν on $\mathbb{R}^N$ and arrive with intensity $\{\lambda(Y_t): t \geq 0\}$, for some $\lambda: D \rightarrow [0,\infty)$, and $\gamma \in \mathbb{R}^K$ is the vector of unknown parameters governing the model for Y_t

- Cox process construction of jumps: conditional on $\{Y_s: 0 \leq s \leq t\}$, the times of jumps in [0, t] are the jump times of a Poisson process with time-varying intensity $\{\lambda(Y_s): 0 \leq s \leq t\}$; the size of the jump distribution ν is assumed to be independent of $\{Y_s: 0 \leq s < T\}$

– The special case of an affine-jump diffusion is obtained by requiring that μ, σσ', and λ all be affine functions on D:

Affine in Y_t

$$dY_t = \mathcal{K}(\Theta - Y_t)\, dt + \Sigma\sqrt{S_t}\, dW_t + dZ_t$$

Potentially non-symm. & non-diagonal

The affine family: Jump diffusion case

$$S_{ii,t} = \alpha_i + \beta_i' Y_t$$ (affine in Y_t)

– When jumps are present, the jump intensity λ(t) is assumed to also be a positive, affine function of the state Y_t

$$\lambda(t) = l_0 + l_Y' Y(t)$$

and the jump-size distribution f_J is assumed to be determined by its characteristic function $\mathcal{J}(u) = \int \exp\{ius\} f_J(s) ds$

– Singleton (2001, JoE) shows that in this case the CCF has a linear affine structure with coefficients satisfying a system or Riccati equations:

$$\mathrm{CCF}_t(\tau, u) = e^{\phi_{0t} + \phi_{Yt}' Y_t}$$

$$\dot{\phi}_{Yt} = \mathcal{K} \cdot \phi_{Yt} - \frac{1}{2}\phi_{Yt}' H_1 \phi_{Yt} - l_Y(\mathcal{J}(\phi_{Yt}) - 1),$$

$$\dot{\phi}_{0t} = -\mathcal{K}\Theta \cdot \phi_{Yt} - \frac{1}{2}\phi_{Yt}' H_0 \phi_{Yt} - l_0(\mathcal{J}(\phi_{Yt}) - 1)$$

where $[\Sigma S_t \Sigma']_{ij} = [H_0]_{ij} + [H_1]_{ij} Y_t$ and boundary conditions $\phi_{YT} = iu$ and $\phi_{0T} = 0$

THE AFFINE FAMILY: JUMP DIFFUSION CASE

– Focusing on case of an affine diffusion (no jumps) and $\beta_i = 0\ \forall i$, the affine structure of the drift and instantaneous variance carries over to the conditional moments of $\{Y_t\}$:

$$E\left[Y_{t+\tau}\middle|Y_t\right] = e^{-\mathcal{K}\tau}Y_t + (I - e^{-\mathcal{K}\tau})\Theta \quad \Omega(\tau) = \left[V_{ij}\frac{1 - e^{-(\kappa_i+\kappa_j)\tau}}{\kappa_i + \kappa_j}\right]_{i,j=1}^{N}$$

$$\mathrm{Var}(Y_{t+\tau}|Y_t) = X^{-1}\Omega(\tau)\left(X^{-1}\right)' \qquad V = X\Sigma^2 X'$$

where X is a NxN matrix s.t. $X\mathcal{K}(X')^{-1}$ is diagonal

- These are obtained evaluating the *n*th derivative of the CCF with respect to u at 0
- $\mathcal{K}$ governs the degree of mean reversion in the process toward its "long-run" or unconditional mean
- This corresponds to a "half-life" of $\exp(-\mathcal{K}\tau) = 0.5$

– Because $\beta_i = 0$ for all i, then Y_t is a Gaussian process

– Alternatively, if $\mathcal{K}$ is diagonal with ith element κ_i , Σ is diagonal

ADMISSIBILITY

with ith element σ_i, β_i is zero except in the ith location (where it is unity), and $\alpha_i = 0$ for all i, then Y is a vector of N independent square root diffusions with conditional variance of Y_i:

$$\sigma_{rt}^2(\Delta) \equiv Var[r_{t+\Delta}|r_t] = r_t \frac{\sigma^2}{\kappa}(e^{-\Delta\kappa} - e^{-2\Delta\kappa}) + \bar{r}\frac{\sigma^2}{2\kappa}(1 - e^{-\Delta\kappa})^2$$

- The fact that CCF_t or $\mathrm{CMGF}_t = e^{\phi_{0t} + \phi_{Yt}' Y_t}$ does not imply that the model is automatically admissible
  - Implicit in the requirements for well defined CCF and CMGF are conditions that ensure that even powered conditional moments of the distribution of Y are nonnegative
  - **Example 6 [Two-factor affine model]**

$$d\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 \\ 0 & \kappa_{22} \end{pmatrix}\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} - \begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix}\right) dt + \begin{pmatrix} \sigma_1\sqrt{Y_{1t}} & 0 \\ 0 & \sqrt{\alpha_2 + \beta_{21} Y_{1t}} \end{pmatrix} dW_t$$

ADMISSIBILITY

- The constraint that Y_{2t} does not appear in the drift and volatility of Y_{1t} imply that the first variable Y_1 is an autonomous square-root diffusion; as such, so long as $\sigma_1 > 0$ and $\kappa_{11}\theta_1 > 0$, Y_1 is guaranteed to be nonnegative
  - For $(Y_1, Y_2)'$ to be well-defined, the instantaneous variances Y_{1t} and $\alpha_2+\beta_{21}Y_{1t}$ must be nonnegative
- Further, this process is specified so that the volatility of Y_{2t} depends only on Y_{1t} : the constraints $\alpha_2 > 0$ and $\beta_{21} > 0$ guarantee that $\alpha_2 + \beta_{21}Y_{1t} \geq 0$
  - This admissible parameterization illustrates the point that if the state vector is divided up into the subvector that drives volatility (Y_{1t}) and the remaining variables (Y_{2t}), and sufficient structure is imposed on the first subvector, then we are assured admissibility up to the imposition of sign restrictions
- There is no admissibility problem if $\beta_i = 0$, for all i, because in

ADMISSIBILITY

the instantaneous conditional volatilities are all constants

– However, outside this special case, to ensure admissibility it is necessary to constrain the drift ($\mathcal{K}$ and Θ) and diffusion $\mathcal{B} \equiv (\beta_1, \beta_2, \ldots, \beta_N)$
– Requirements for admissibility become increasingly stringent as the number of state variables determining $S_{ii,t}$ increases
– To formalize this consider the case where there are M state variables (WLOG, the first M) driving the instantaneous conditional variances of the N-vector Y, so M = rank($\mathcal{B}$)
– We define a set of N+1 benchmark models $\mathbb{A}_M(N)$ as follows: (1) partition Y_t as $Y \equiv (Y^{V\prime}, Y^{D\prime})$, where Y^V is M × 1 and Y^D is (N−M)×1, and define

$$\mathcal{K} = \begin{bmatrix} \mathcal{K}^{VV}_{M\times M} & 0_{M\times(N-M)} \\ \mathcal{K}^{DV}_{(N-M)\times M} & \mathcal{K}^{DD}_{(N-M)\times(N-M)} \end{bmatrix}$$

($\mathcal{K}$ unconstrained if M=0)

ADMISSIBILITY

Diagonal matrix with elements σ_i, i=1,2,...

(2)

$$\Sigma = \begin{bmatrix} \sigma I_{M\times M} & 0_{M\times(N-M)} \\ \Sigma^{DV}_{(N-M)\times M} & \Sigma^{DD}_{(N-M)\times(N-M)} \end{bmatrix}$$

(3)

$$\alpha = \begin{bmatrix} 0_{M\times 1} \\ \alpha^{D}_{(N-M)\times 1} \end{bmatrix} \geq 0 \quad \mathcal{B} = \begin{bmatrix} I_{M\times M} & B^{VD}_{M\times(N-M)} \\ 0_{(N-M)\times M} & 0_{(N-M)\times(N-M)} \end{bmatrix}$$

Rank($\mathcal{B}$) = M

(4) under the following parametric restrictions:

$$\mathcal{K}_i\Theta \equiv \sum_{j=1}^{M} \mathcal{K}_{ij}\Theta_j > 0, \ 1 \leq i \leq M$$

$$\mathcal{K}_{ij} \leq 0, \ 1 \leq j \leq M, \ j \neq i,$$

$$\mathcal{B}_{ij} \geq 0, \ 1 \leq i \leq M, \ M+1 \leq j \leq N$$

– $\mathbb{A}_M(N)$ implies that the conditional variances are controlled by the first M state variables

$$S_{ii,t} = Y_{it}, \quad 1 \leq i \leq M, \ S_{jj,t} = \alpha_j + \sum_{k=1}^{M} \beta_{jk} Y_{kt}, \quad M+1 \leq j \leq N$$

ADMISSIBILITY

- As long as $Y^V_t \equiv (Y_1, Y_2, ..., Y_M)$ is nonnegative with probability one, the benchmark representation of $Y_t \equiv (Y^V_t, Y^D_t)$ is admissible
- To analyze this, note that Y^V_t follows the diffusion

$$dY_t^V = (\mathcal{K}^{VV}\ \mathcal{K}^{VD})(\Theta - Y_t)dt + (\Sigma^{VV}\ \Sigma^{VD})\sqrt{S_t}\,dW_t$$

- To ensure that Y^V_t is bounded at zero from below, the drift of Y^V_t must be nonnegative and its diffusion must vanish at the zero boundary. Necessary and sufficient conditions for this are C1:

$$\text{C1: } \mathcal{K}^{VD} = 0_{M\times(N-M)};\ \text{C2: } \Sigma^{VD} = 0_{M\times(N-M)};$$

$$\text{C3: } \Sigma_{ij} = 0, 1 \leq i \neq j \leq M;\ \text{C4: } \mathcal{K}_{ij} \leq 0, 1 \leq i \neq j \leq M;$$

$$\text{C5: } \mathcal{K}^{VV}\Theta^V > 0$$

  - Please read Singleton as to the specific reasons for each of C1-C5
- Notice that not all affine models are special cases of these canonical models. For models outside these subfamilies, admissibility should be verified, on a case-by-case basis

DISCRETE TIME AFFINE MODELS

- General principle: a discretization of a continuous-time model typically does not lead to a well-defined discrete-time counterpart of an $\mathbb{A}_M(N)$ model
- To construct discrete-time affine models with same features of the $\mathbb{A}_M(N)$ family one is better off starting from the primitive assumption that the CMGF of Y_{t+1} is an exponential-affine function of Y_t, $\text{CCF}_t \text{ or } \text{CMGF}_t = e^{\phi_{0t}+\phi'_{Y_t} Y_t}$
- Importantly, starting with the CMGF may allow for richer formulations of the dynamics of Y than in standard affine diffusion models
- A discrete-time affine process is obtained by positing a functional form for the ϕ_{0t} and ϕ_{Yt} that defines CMGF_t
  - E.g ., $Y_{t+1}|Y_t \sim N(\alpha+\beta Y_t, \sigma^2) \Rightarrow \text{CMGF}_t(u) = e^{u(\alpha+\beta Y_t)+u^2\sigma^2/2}$
  - This can be interpreted as the discrete-time counterpart to

DISCRETE TIME AFFINE MODELS

continuous-time models as length of the sampling interval shrinks to zero:

- Let $\alpha = \kappa\theta\Delta t$, $\beta = 1 - \kappa\Delta t$, and $\sigma^2 = \tilde{\sigma}^2\Delta t$, then as Δt → 0, DT process converges to Gaussian O-U: $dY_t = \kappa(\theta - Y_t)dt + \tilde{\sigma}\, dW_t$

– Another popular case occurs under the assumption that the CMGF of a scalar Markov process Y is given by

$$E_t\left[e^{uY_{t+1}}\right] = e^{-\nu \ln(1-uc)-(\rho u/1-uc)Y_t} \quad \rho > 0, c > 0, \text{and } \nu > 0$$

– This is the CMGF of an autoregressive gamma (AG) process obtained from setting

$$Y_{t+1}|(Z_{t+1}, Y_t) \sim c \text{ gamma } (\nu + Z_{t+1})$$
$$Z_{t+1}|Y_t \sim \text{ Poisson } (\rho Y_t/c)$$

- First two conditional moments of Y_{t+1}:
$$E\left[Y_{t+1}\middle|Y_t\right] = \nu c + \rho Y_t$$
$$\text{Var}\left[Y_{t+1}\middle|Y_t\right] = \nu c^2 + 2c\rho Y_t$$

- Also in this case, the continuous time limit of the process, based

Discrete Time Affine Models

on $\rho = 1 - \kappa\Delta t,\ c = (\sigma^2/2)\Delta t,\ \nu = 2\kappa\theta/\sigma^2$ shows that

$$\frac{E_t\left[Y_{t+\Delta t}\right] - Y_t}{\Delta t} \to \kappa(\theta - Y_t), \quad \frac{\mathrm{Var}_t\left[Y_{t+\Delta t}\right]}{\Delta t} \to \sigma^2 Y_t$$

so that Y_t converges to square-root process:

$$dY_t = \kappa(\theta - Y_t)dt + \sigma\sqrt{Y_t}dW_t$$

- To construct the discrete-time counterparts to many multivariate affine diffusions, we need to combine multivariate Gaussian and AG processes
- In a series of papers, Dai, Le and Singleton have partioned the Nx1 vector $X_t \equiv (Z'_t, Y'_t)'$ and defined it as

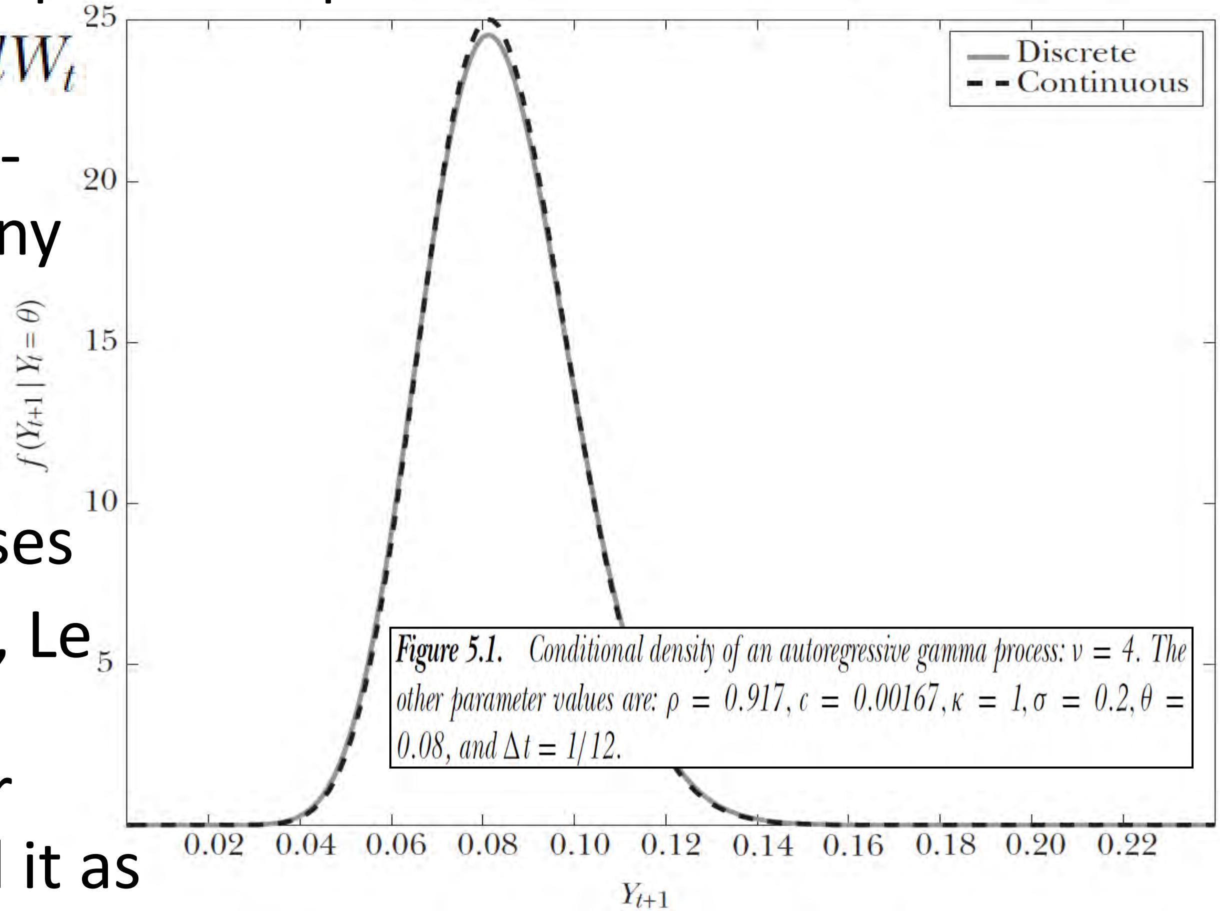


Figure 5.1. Conditional density of an autoregressive gamma process: $\nu = 4$. The other parameter values are: $\rho = 0.917$, $c = 0.00167$, $\kappa = 1$, $\sigma = 0.2$, $\theta = 0.08$, and $\Delta t = 1/12$.

DISCRETE TIME AFFINE MODELS

a $D\mathbb{A}_M(N)$ process if: (i) Z_t is an autonomous $D\mathbb{A}_M(M)$ process; and (ii) conditional on Y_t and Z_t, Y_{t+1} is normally distributed with a conditional variance that depends on Z_t; (iii) the $M \times 1$ autonomous Z_t follows a vector autoregressive gamma process

- To construct a joint density of X_{t+1} given X_t, they exploit $f(Z_{t+1}|X_t) = f(Z_{t+1}|Z_t)$ and that, Z_{t+1} and Y_{t+1} are independent

– This assumption amounts to within-period shocks to Z and Y being independent, which is more restrictive than for continuous-time models since the latter models allowed for nonzero (instantaneous) correlations across the Y^V and Y^D models

– Under these assumptions, the conditional Laplace transform of X_{t+1} given X_t is exponential-affine and in the continuous time limit, X_t becomes the canonical model $\mathbb{A}_M(N)$

- As seen, Fourier transforms of the conditional density (the CCF) of Y_{t+1}, discretely sampled from an affine diffusion, is

TRANSFORMS FOR AFFINE MODELS

known in closed form, CCF_t = $e^{\phi_{0t}+\phi'_{Yt}Y_t}$; it turns out that more general transforms of affine processes are known and these transforms play a central role in pricing bonds

- Start by introducing the discount rate function R: D → $\mathbb{R}$ and assume it is also affine $R(Y) = \rho_0 + \rho_1 \cdot Y$
- A transform is based on a "characteristic" $\chi = (K, H, l, \theta, \rho)$ where ρ are the parameters of the discount function and K, H, l, and θ are parameters of the affine model:

$$\mathbb{T}^\chi : \mathbb{C}^N \times D \times \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{C} \text{ of } Y_T \text{ conditional on } \mathcal{F}_t$$

$$\mathbb{T}^\chi(u, Y_t, t, T) = E^\chi\left(\exp\left(-\int_t^T R(Y_s)ds\right) e^{u\cdot Y_T} \,\middle|\, \mathcal{F}_t\right)$$

  where E^χ is under the distribution of Y determined by χ

  - T^χ differs from the CCF of the distribution of Y_T because of the discounting at rate $R(Y_t)$

TRANSFORMS FOR AFFINE MODELS

– Duffie, Pan, and Singleton (2000, ECMA) show that under technical regularity conditions:

$$\mathbb{T}^{\chi}(u, y, t, T) = e^{\psi_{0t} + \psi_{Yt} \cdot y}$$

where ψ_{0t} and ψ_{Yt} satisfy a system of complex valued ODEs

- In some applications, explicit solutions for these ODEs are known (e.g., for square-root diffusions); in other cases, solutions are found numerically

- How do you estimate affine models? First idea: because differentiating the CCF yields conditional moments, then GMM is a natural, albeit inefficient, approach

– These moments can be computed from the derivatives of the CCF (or CMGF when it exists) evaluated at u = 0:

$$\frac{\partial^{j+k} \mathrm{CCF}_t(u, \gamma_0)}{\partial u_{s_1}^{j} u_{s_2}^{k}} \Big|_{u=0} = i^{j+k} E\left[Y_{s_1,t+1}^{j} Y_{s_2,t+1}^{k} \,\middle|\, Y_t\right]$$

GMM AND QML ESTIMATION

so that the orthogonality conditions are

$$E\left(Y^{j}_{s_1,t+1} Y^{k}_{s_2,t+1} - \frac{\partial^{j+k} \mathrm{CCF}_t(u,\gamma_0)}{i^{j+k} \partial u^{j}_{s_1} u^{k}_{s_2}} \Bigg|_{u=0} \Bigg| \, Y_t \right) = 0$$

– Although in a GMM it would be a bad idea to only use the first two moments, with the first two conditional moments in hand, a QML estimator with the normal log-lik can be computed

– However, both GMM and QML are dramatically inefficient in the case of affine models because—apart from the case of Gaussian diffusions—the "innovations" in affine models are nonnormal

- E.g., noncentral chi-square in the case of square-root diffusions

– The "high way" in the case of affine models is represented by methods that exploit (at least to some extent) knowledge of the density functions

- Three options: (a) full-information ML; (b) simulated ML;

MAXIMUM LIKELIHOOD ESTIMATION

(3) approximate MLE

- If the conditional density of Y_{t+1} given Y_t , $f(Y_{t+1}|Y_t; \gamma_0)$, is known in closed form, then we can proceed directly to write and maximize the conditional log-likelihood of the sample
  - Examples of affine diffusions with known conditional density functions are the cases where Y_t is a vector Gaussian process (M = 0) and Y_t is a vector of independent square-root processes (M = N and $\mathcal{K}$ is diagonal)
  - In the former case, $Y_{t+\tau}$ conditional on Y_t is normally distributed, in the latter case noncentral chi-square
- An example of a jump-diffusion model with a known likelihood function is the pure-jump diffusion

$$dY(t) = \mu_Y dt + \sigma_Y \, dW(t) + J(t) \, dZ(t)$$

where the jump amplitude J(t) is distributed as $N(m_J, \delta^2)$

- Conditional on the number of jumps $L = \ell$ (over the sampling

Maximum Likelihood Estimation

interval of the data), the distribution of ΔY_t is

$$f(\Delta Y_t | L = \ell) = N(\mu_Y + \ell\zeta, \eta^2 + \ell\delta^2)$$

- This is an implication of the jump amplitudes J_t be normally distributed and that, conditional on L, ΔY_t is the sum of L + 1 independent normal random variables, N $(n\mu_Y, n\sigma^2_Y)$ + J_1 + ... + J_L
- The density of ΔY_t is then obtained by "integrating out" over the density of the number of jumps:

$$f(\Delta Y_t) = \sum_{\ell=0}^{\infty} \frac{e^{-\lambda}\lambda^{\ell}}{\ell!} \frac{1}{\sqrt{2\pi(\eta^2 + \ell\delta^2)}} e^{-\frac{1}{2}(\Delta Y_t - \mu_Y - \ell\psi)^2/(\eta^2 + \ell\delta^2)}$$

- One advantage of casting an affine model in discrete time is that the likelihood of the data is known in closed form for a much larger class of models than in continuous case
  – If Z_{t+1} conditional on Z_t follows an autoregressive gamma process, Y_{t+1} conditional on $X_t = (Z_t, Y_t)$ is Gaussian, then the

SIMULATED MAXIMUM LIKELIHOOD

likelihood function for X can be constructed in closed form

- The idea is to divide each sampling interval [t, t+1] into n subintervals, say of equal length *h* = (1/*n*), and expressing the density function of the data as

$$f(Y_{t+1}|Y_t) = \int_{\mathbb{R}^N} f(Y_{t+1}|Y_{t+1-h}) \times f(Y_{t+1-h}|Y_t)\, dY_{t+1-h}$$
$$= E\big[f(Y_{t+1}|Y_{t+1-h};\gamma)\big|Y_t;\gamma\big]. \qquad (*)$$

- The density function $f(Y_{t+1}|Y_t)$ can be interpreted as an expectation of $f(Y_{t+1}|Y_{t+1-h})$, treated as a function of Y_{t+1-h}, and integrated against the conditional density $f(Y_{t+1-h}|Y_t)$
- Therefore, if the density $f(Y_{t+1}|Y_{t+1-h})$ can be accurately approximated, and (given Y_t) Y_{t+1-h} can be simulated, then (*) can be computed by Monte Carlo integration
  - Pedersen (1995, SJS) proposes replacing the density $f(Y_{t+1}|Y_{t+1-h})$ in

Simulated Maximum Likelihood

(*) with $f_N(Y_{t+1} | Y^{\gamma,n}_{t+1-h})$, the density fnct. of a normal distribution with mean $(1/n)\mu(Y^{\gamma,n}_{t+1-h}; \gamma)$ and variance $\sigma^2(Y^{\gamma,n}_{t+1-h}; \gamma)/n$

– This approach relies on Euler approximations multiple times to simulate $Y^{\gamma,n}_{t+1-h}$ given Y_t; therefore, in establishing the large-sample properties, the nature of the approximation errors have to be examined with limiting distributions of sample moments

– Pedersen shows, for the special case of a Gaussian diffusion, that there is a rate at which *n* can grow with T such that consistency and asymptotic normality are ensured

  - The approximation errors approach zero at a sufficiently fast rate relative to T that these errors can be ignored in the computation of the asymptotic distribution of the ML estimator for γ_0

– Brandt and Santa-Clara (2002, JFE) show that, if $T \rightarrow \infty$, $n \rightarrow \infty$, and $\mathcal{T} \rightarrow \infty$, with $\mathcal{T}^{1/2}/n \rightarrow 1$ and $T/\mathcal{T}^{1/4} \rightarrow 0$, then their simulated ML estimator is asymptotically normal

SIMULATED MAXIMUM LIKELIHOOD

- The literature on SMLE is constantly growing, especially in the direction of exploiting the specific structure of affine models, e.g., jump diffusion model
- For instance, Duffie, Pedersen and Singleton (2003, JF) stress that many affine models contain autonomous components, e.g.,

$$dY_1 = (k_1 - K_{11}Y_1)dt + \sqrt{Y_1}\, dW_1,$$

$$dY_2 = (k_2 - K_{21}Y_1 - K_{22}Y_2)dt + \sqrt{1 + \beta Y_1}\, dW_2$$

  that can be exploited, like in the case

$$f(Y_t | Y_{t-h}) = f(Y_{1t} | Y_{1,t-h}) \times f(Y_{2t} | Y_{1t}, Y_{t-h})$$

- The structure of the model is such that f $(Y_{1t}|Y_{1,t-h})$ is known exactly to be a noncentral chi-square distribution; no approximation for its conditional density function is necessary, so that Pedersen's Euler approximations are unnecessary

Approximate MLE

- The approach uses much more information than the one developed by Pedersen (obviously, this is true with regard to Y_1, since the noncentral chi-square is known exactly)

- An alternative approach to ML estimation, based on polynomial approx., was proposed by Ait-Sahala (2002, JoE)
  - For instance, suppose that Y follows a univariate diffusion with drift $\mu_Y(Y_t; \gamma)$ and instantaneous volatility $\sigma_Y(Y_t ; \gamma)$
  - Ait-Sahalia's approximation begins by transforming Y to have unit volatility by means of the transformation

$$X_t \equiv \psi(Y_t;\gamma) = \int^{Y_t} \frac{du}{\sigma_Y(u;\gamma)} \overset{\text{Ito's lemma}}{\Rightarrow} dX_t = \mu_X(X_t;\gamma)dt + dW_t$$

  - The basic idea then is to approximate the logarithm of the conditional density of X using Hermite polynomials, and then to use a change of-variable to obtain the log-density of Y:

APPROXIMATE MLE

– Specifically, letting Δ denote the time interval between discrete observations and assuming an expansion out to order K in powers of Δ gives:

$$\ln f(\Delta, x|X_0) = \underbrace{-\frac{\ln(2\pi\Delta)}{2} - \frac{1}{2\Delta}(x - X_0)^2}_{\text{Density of N(0,1)}} + \sum_{k=0}^{K} C_X^{(k)}(x|X_0;\gamma)\frac{\Delta^k}{k!}$$

where the $C_X^{(k)}$ are constructed recursively from integrals of μ_X and its derivatives

– The expansion of the conditional density of Y is obtained by a change of variable ($\nabla\psi(y;\gamma) = \sigma_Y^{-1}(y;\gamma)$):

$$f_Y(\Delta, y|Y_0;\gamma) = \operatorname{Det}[\nabla\psi(y;\gamma)] f_X(\Delta, \psi(y;\gamma)|\psi(Y_0;\gamma);\gamma)$$

– Taking logarithms gives the approximate log-likelihood function:

$$\ln f_Y(\Delta, y|Y_0;\gamma) = -\frac{\ln(2\pi\Delta)}{2} + \ln \operatorname{Det}[\nabla\psi(y;\gamma)] +$$

CHARACTERISTIC FUNCTION-BASED ESTIMATION

$$-\frac{1}{2\Delta}(\psi(y;\gamma)-\psi(Y_0;\gamma))^2+\sum_{k=0}^{K} C_X^{(k)}(\psi(y;\gamma)|\psi(Y_0;\gamma))\frac{\Delta^k}{k!}$$

- Given that the CCF and/or CMGF of an affine process is known (at least up to the solution of ODEs), estimation can be based on these representations of the conditional distribution even when functional form is unknown
  - Since the functional form of the CCF of an affine process is known, the conditional density function of Y_{t+1} is also known up to an inverse Fourier transform of the CCF_t:

$$f_Y\left(Y_{t+1}\middle|Y_t;\gamma\right)=\frac{1}{(2\pi)^N}\int_{\mathbb{R}^N} e^{-iu'Y_{t+1}}\mathrm{CCF}_t(u,\gamma)\,du$$

which gives the implicitly defined log-likelihood function:

$$l_T(\gamma)=\frac{1}{T}\sum_{t=1}^{T}\log\left\{\frac{1}{(2\pi)^N}\int_{\mathbb{R}^N} e^{-iu'Y_{t+1}}\mathrm{CCF}_Y(u,\gamma)\,du\right\}$$

CHARACTERISTIC FUNCTIONS

Definition: If X is a (single) random variable, then its characteristic function $c_X: \mathbb{R} \to \mathbb{C}$ is defined by

$$c_X(\xi) \equiv \mathbb{E}\left(e^{i\xi X}\right), \xi \in \mathbb{R} \text{ arbitrary} \qquad (1)$$

If X has a probability density, $f_X = f_X(x)$, then this becomes

$$c_X(\xi) = \int_{\mathbb{R}} e^{i\xi x} f_X(x) dx \qquad (2)$$

Remark: The RHS of (2) with ξ replaced by $(-\xi)$ is known as Fourier transform of f_X

CHARACTERISTIC FUNCTIONS

- **Fourier Inversion**

If X has characteristic function $c_X(t)$, then for any interval (a, b),

$$P[a < X < b] + \frac{P[X = a] + P[X = b]}{2} = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} c_X(t) dt$$

- **Conditional Characteristic Function**

of X_2 knowing $X_1 = x_1$

$$c_{X_2|X_1=x_1}(\xi) \equiv \mathbb{E}\left(e^{i\xi X_2} | X_1 = x_1\right) = \int_{\mathbb{R}} e^{i\xi x_2} f_{X_2|X_1}(x_2|x_1) dx_2$$

Characteristic function-based estimators

- If the Conditional Characteristic Function (CCF) and/or the *Conditional Moment Generating Function* (CMGF) of an affine process is known, estimation can be based on these representations
- In case of $\tau = 1$ the CCF of Y_{t+1} on Y_t can be represented as $\mathrm{CCF_t(u,\gamma)}$, where γ is the unknown parameter vector and γ_0 denotes true population value of γ

ML ESTIMATION BY FOURIER INVERSION

The conditional density function of Y_{t+1} is known up to an Inverse Fourier transform of $CCF_t(u,\gamma)$:

$$f_Y(Y_{t+1}|Y_t;\gamma) = \frac{1}{(2\pi)^N}\int_{\mathbb{R}^N} e^{-iu'Y_{t+1}} CCF_t(u,\gamma)du$$

Then the log-likelihood function $l_T(\gamma)$ is:

$$l_T(\gamma) = \frac{1}{T}\sum_{t=1}^{T} \log\left\{\frac{1}{(2\pi)^N}\int_{\mathbb{R}^N} e^{-iu'Y_{t+1}} CCF_t(u,\gamma)du\right\}$$

The resulting estimator is called ML-CCF estimator

ML-CCF ESTIMATOR, CIR EXAMPLE (SINGLETON, 2001)

Conditional distribution of r_{t+1} on r_t is non central χ^2, then CCF for r_{t+1} is:

$$CCF_t(u) = (1 - iu/c)^{-(2\kappa\theta/\sigma^2)} exp\left\{\frac{iue^{-\kappa}r_t}{(1 - iu/c)}\right\}$$

The conditional density of r_{t+1} is computed by Gauss-Legendre quadrature. With 20 quadrature points the ML-CCF estimator is identical to the computed using non central χ^2 distribution of r_{t+1}

Weekly data on a short term interest rate series, average annualized short rate is about 6.0%.

	κ	θ	σ
Population	0.4	6.0	0.3
$q_p = 10$	0.889	5.930	0.303
	(0.32)	(0.19)	(0.007)
$q_p = 20$	0.377	5.621	0.302
	(0.20)	(0.46)	(0.007)
$q_p = 50$	0.377	5.621	0.302
	(0.20)	(0.46)	(0.007)

ML-CCF

additionally

Partial ML-CCF Estimator

Problem: $(q_p)^N$ increase in the number of points in the grid of approximation in the Fourier Inversion, where N is the dimension of Y_t and q_p - number of quadrature points.

Solution: using conditional density function of the individual elements of Y:

$$f_j(Y_{j,t+1}|Y_t;\gamma) = \frac{1}{(2\pi)}\int_{\mathbb{R}} e^{-i\omega\iota_j' Y_{t+1}} CCF_t(\omega\iota_j,\gamma)d\omega$$

where $Y_{j,t+1} = \iota_j \cdot Y_{t+1}$ and ι_j N-dimensional vector s.t.

$$\iota_j = \begin{cases} 1\ at\ jth\ position \\ 0\ o/w \end{cases}$$

Partial ML-CCF Estimator

- For fixed j GMM estimator:

$$E\left[\frac{\partial \log f_j}{\partial \gamma}(Y_{j,t+1}|Y_t,\gamma_0)\right] = 0$$

- Objective function:

$$l_{jT}(\gamma) = \frac{1}{T}\sum_{t=1}^{T}\log\left\{\frac{1}{(2\pi)}\int_{\mathbb{R}} e^{-i\omega\iota_j{}'Y_{t+1}} CCF_t(\omega\iota_j,\gamma)d\omega\right\}$$

- FOCs:

$$\frac{\partial l_{jT}}{\partial \gamma}(\gamma_T) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{f_j(Y_{j,t+1}|Y_t,\gamma_T)} \times \frac{1}{(2\pi)}\int_{\mathbb{R}} e^{-i\omega\iota_j{}'Y_{t+1}} CCF_t(\omega\iota_j,\gamma)d\omega = 0$$

γ_T is a consistent and asymptotically normal GMM estimator of γ_0

FREQUENCY-DOMAIN ESTIMATOR

ECCF Estimator

- ECCF – empirical conditional characteristic function
- Set Z_T^{∞} of "instrument" functions with elements $z_t(u)$: $\mathbb{R}^N \to \mathbb{C}^Q$ with $z_t(u) \in \mathcal{F}_t$, $z_t(u) = \bar{z}_t(-u)$, $t = 1, \dots, T$ and $\mathcal{F}_t$ is the σ-algebra generated by Y_t
- *"Residual"*: $\epsilon_{t+1}(u,\gamma) \equiv e^{iu'Y_{t+1}} - CCF_t(u,\gamma)$, where $CCF_t(u,\gamma) = E\left[e^{iu'Y_{t+1}}|Y_t\right]$
- Each $z \in Z_T^{\infty}$ indexes an estimator $\gamma_{\infty T}^{z}$ of γ_0 satisfying

$$\frac{1}{T}\sum_t \int_{\mathbb{R}^N} z_t(u)\, \epsilon_{t+1}(u, \gamma_{\infty T}^{z})du = 0 \quad (3)$$

FREQUENCY-DOMAIN ESTIMATOR

- $\gamma^{z}_{\infty T}$ is consistent, and asymptotically normal with limiting covariance matrix:

$$\mathcal{V}_0^{\infty}(z) = D(z)^{-1}\Sigma^{\infty}(\bar{D}(z)')^{-1},$$

where

$$D(z) = E\left[\int_{\mathbb{R}^N} z_t(u)\frac{\partial \phi_t(u)}{\partial \gamma}du\right]$$

$$\Sigma^{\infty}(z) = E\left[\int_{\mathbb{R}^N} z_t(u)\epsilon_{t+1}(u,\gamma_0)du\int_{\mathbb{R}^N}\bar{\epsilon}_{t+1}(u,\gamma_0)\bar{z}_t(u)'du\right]$$

- Optimal index (the smallest asymptotic covariance matrix among empirical CCF estimators):

$$z^{*}_{\infty t}(u) = \frac{1}{(2\pi)^N}\int_{\mathbb{R}^N}\frac{\partial \log f}{\partial \gamma}(Y_{t+1}|Y_t,\gamma_0)'e^{-iu'Y_{t+1}}dY_{t+1}$$

GMM-CFF Estimator

- $z^*_{\infty t}$ cannot be computed without a prior knowledge of the conditional density function
- Approximate the integral (Y_t is one dimensional):

$$\int_{\mathbb{R}} z_t(u) \left[e^{iu'Y_{t+1}} - CCF_t(u,\gamma)\right] du$$

with the sum over a finite grid in $\mathbb{R}$ and substitute into (3)

- GMM-CCF estimator is consistent and asymptotically normal
- As the range and fineness of the approximating grid $\mathbb{R}$ increases asymptotic covariance matrix of optimal GMM-CCF estimator converges to $I(\gamma_0)^{-1}$ - asymptotic Cramer-Rao lower bound.

NON-AFFINE MODELS: HINTS

- The distributions of asset returns are "fat tailed" and "skewed"; this can be due to "stochastic volatility" (including ARCH) and "jumps" (including regimes)
- The (unconditional) skewness of a random variable *r*, defined as $\text{Skew} = E[(r - E(r))^3]/\sigma^3$ is measure of asymmetry
- Kurtosis is measure of tail-fatness, $\text{Kurt} = E[(r - E(r))^4]/\sigma^4$
  - The kurtosis of a normal random variable is 3

	Equity returns (daily)									
	Aust	Can	Fra	Ger	HK	Jap	Spain	Swe	UK	US
μ	0.019	0.024	0.021	−0.022	0.030	−0.022	0.030	0.035	0.017	0.032
σ	0.858	0.963	1.296	0.922	1.640	1.307	1.360	1.555	1.026	1.025
Skew	−0.237	−0.547	−0.134	0.549	−0.033	0.233	−0.157	0.170	−0.107	−0.114
Kurt	5.952	10.787	5.871	6.655	11.613	6.905	6.459	6.720	6.160	6.857

	Swap holding-period returns (daily)									
	Aust	Can	Fra	Ger	HK	Jap	Spain	Swe	UK	US
μ	0.035	0.034	0.030	0.028	0.025	0.019	0.041	0.037	0.038	0.030
σ	0.391	0.320	0.227	0.205	0.427	0.195	0.316	0.351	0.259	0.280
Skew	−0.306	−0.355	−0.025	−0.165	−0.547	−0.256	−0.075	0.009	0.580	−0.273
Kurt	8.180	8.232	6.303	6.817	28.270	8.138	17.144	23.319	14.263	5.298

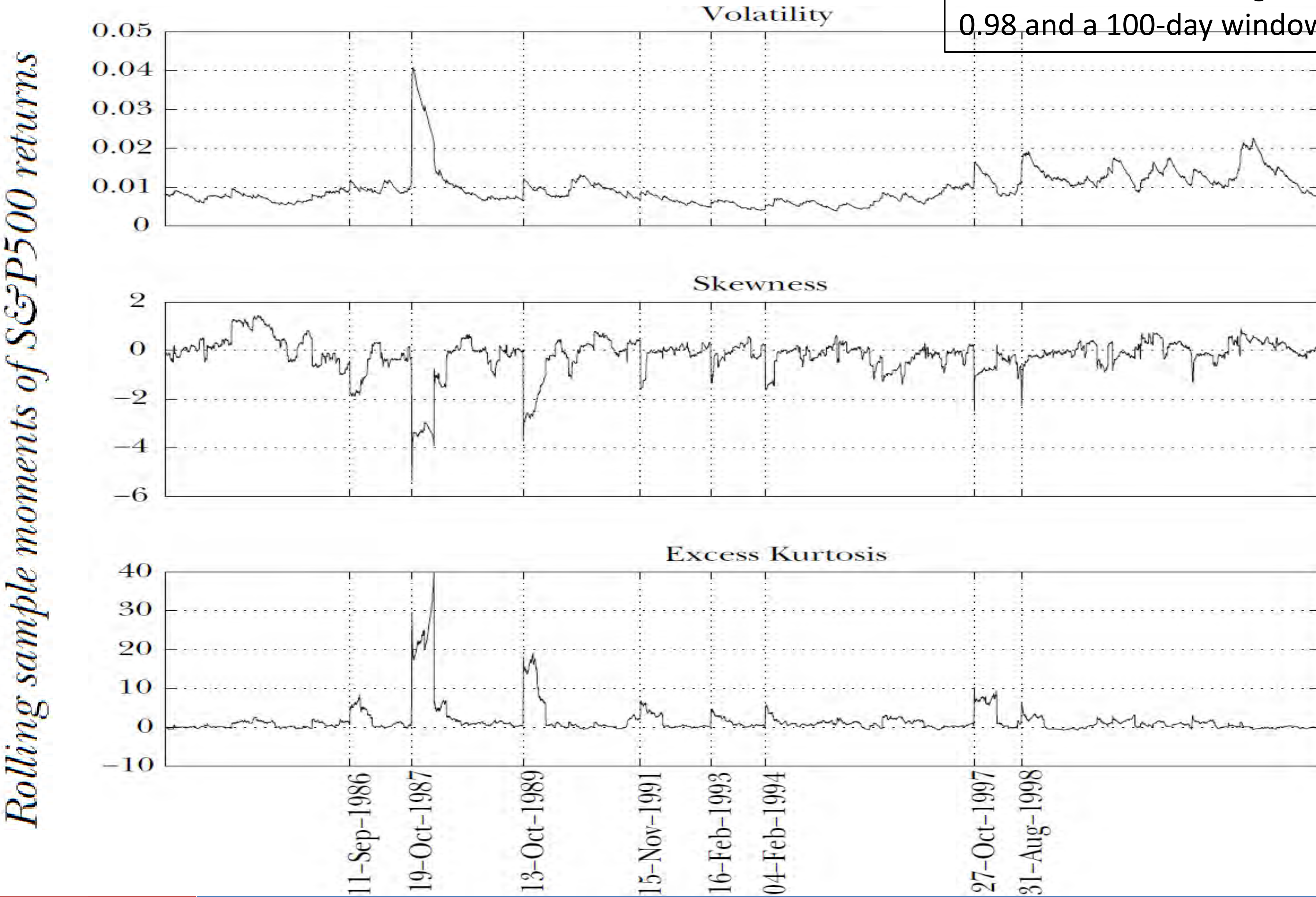
NON-AFFINE MODELS: HINTS

- Holding period returns exhibit excess kurtosis or fat tails: whatever the econometric model of returns adopted, it should imply nonnormal, fat-tailed marginal return distributions
  - The nonnormality of the marginal distributions of returns need not imply the nonnormality of the conditional distributions
- One simple "probability model" that characterizes dependence of conditional moments on market conditions says that conditional moments of returns are approximated by computing rolling historical sample moments over a fixed window of data, possibly with declining weighting of past observations
- Standardizing returns does not change the picture of strongly time-varying volatility and higher order moments

- We focus on a small number of (especially, discrete) time series models that capture some of the key features of the most commonly used ones

NON-AFFINE MODELS: HINTS

Geometric weighting of past observations with weight factor 0.98 and a 100-day window



NON-AFFINE MODELS: GARCH

– A widely studied formulation of stochastic volatility is the discrete -time GARCH(p,q):

$$r_t = \mu_{t-1} + \sigma_{t-1}\epsilon_t$$

Based on time t-1 information

$$\sigma^2_{t-1} = \omega + \sum_{j=1}^{p} \alpha_j (r_{t-j} - \mu_{t-j-1})^2 + \sum_{i=1}^{q} \beta_i \sigma^2_{t-i-1}$$

where μ_{t-1} is the mean of r_t conditioned on the history of returns at date t − 1, $\{\epsilon_t\}$ is a sequence of i.i.d. N (0, 1) shocks, and ω, (α_j : j = 1, . . . , p), and (β_i : i = 1, . . . , q) are nonnegative

– One measure of persistence is the degree of autocorrelation of σ^2_t : within the GARCH(1,1) model, the coefficient on σ^2_{t-1} in the projection $E[\sigma^2_t | I_{t-1}]$ is α+β

- The condition (α+β)<1 ensures covariance stationarity

– An alternative measure of persistence is the median lag of past $(r_{t-j} - \mu_{t-j-1})^2$ in the conditional variance expression itself

Non-affine models: GARCH

– Writing the conditional variance in a GARCH model as

$$\sigma_t^2 = \bar{\omega} + \sum_{j=0}^{\infty} \delta_j \left(r_{t-j} - \mu_{t-j-1}\right)^2$$

define median lag as ν that satisfies $\left(\sum_{j=0}^{\nu-1} \delta_j\right) / \left(\sum_{j=0}^{\infty} \delta_j\right) = \frac{1}{2}$

– For the GARCH(1,1) model $\nu = -\log 2/\log \beta$, which depends only on β as it is this parameter that governs the rate of geometric decay of the effect of past squared return shocks on σ^2_t

– The dependence of σ^2_t on ϵ^2_t in all of these GARCH-style models implies that return shocks have a symmetric effect on volatility
  - A large positive or negative return shock of equal magnitude in absolute value has the same effect on volatility

– For many markets, and in particular many equity markets, it has long been recognized that positive and negative shocks have asymmetric effects (with leverage) on volatility

Non-affine models: EGARCH & GJR TARCH

- Large negative shocks have a larger effect than correspondingly large positive shocks

– Motivated by the evidence of asymmetry, several researchers have proposed "asymmetric" GARCH-like models, such as Nelson's (1991) logarithmic specification (EGARCH):

$$\ln \sigma_t^2 = \omega + \alpha(|\epsilon_t| - E|\epsilon_t| + \beta\epsilon_t)$$

- Unlike GARCH, with $\beta \neq 0$, the effect of ϵ_t on σ^2_t is asymmetric: if $-1 < \beta < 0$, then $\epsilon_t < 0$ has a larger (positive) effect on volatility than a positive return surprise of the same absolute magnitude

– Modifications to allow for asymmetry have been also proposed by Glosten et al. (1993) and Heston and Nandi (2000)

$$\sigma_t^2 = \omega + \alpha\epsilon_t^2 + \gamma\epsilon_t^2 1_{\{\epsilon_t \geq 0\}} + \beta\sigma_{t-1}^2$$

– With $\gamma < 0$, positive return shocks increase volatility less than negative shocks inducing asymmetry

NON-AFFINE MODELS: NAGARCH & SV

- Alternatively, the NAGARCH model posits

$$r_t = \mu + \lambda\sigma_{t-1}^2 + \sigma_{t-1}\epsilon_t, \qquad \epsilon_t \sim N(0,1)$$

$$\sigma_t^2 = \omega + \alpha\left(\epsilon_t - \gamma\sigma_{t-1}\right)^2 + \beta\sigma_{t-1}^2$$

  - Hence, with $\gamma > 0$, a large negative return shock raises σ^2 more than a large positive shock
- Though volatility is time varying in GARCH, there is no source of randomness to volatility over and above past return shocks

- One can allow for "true" stochastic volatility in discrete-time models by introducing a random volatility shock
  - E.g., following Taylor (1986) assume $\ln\sigma_t^2 = \omega + \beta\ln\sigma_{t-1}^2 + \sigma_v\nu_t$
  - Alternatively, we could assume that σ^2_t follows an auto-regressive gamma AG(a, bσ^2_{t-1}, c) process:

$$\sigma_t^2 = ca + b\sigma_{t-1}^2 + \eta_{v,t-1}\nu_t \qquad \eta_{vt}^2 = c^2a + 2bc^2\sigma_{t-1}^2$$

NON-AFFINE MODELS: MARGINAL DENSITIES

where $\nu_t \sim (0, 1)$, independent of return shocks

– As we have seen, a common assumption in SV volatility models for returns is that ϵ is Gaussian
  - The conditional distribution of returns is a normal
– However, for many financial markets, the distributions of standardized returns $(r_t - \mu_{t-1})/\sigma_{t-1}$ exhibit substantial excess kurtosis and non-zero skewness
– To match the higher-order moments of returns, we can extend the model by introducing fat-tailed shocks to returns in 3 ways:
  - (A) ϵ drawn from a fatter-tailed distribution than a normal;
  - (B) allow the conditional distribution of *r* to possibly change over time, with switching governed by a regime process;
  - (C) Jumps
– As for (A), Bollerslev (1987, JoE) assumes a t distribution and

NON-AFFINE MODELS: MIXTURES & JUMPS

Baille & Bollerslev (1989, JBES) use power exponential dstr.

– As for the mixture distribution case, assume that ϵ follows:

$$\epsilon_t \sim \begin{cases} N(0, \sigma_1^2) & \text{with probability } p \\ N(0, \sigma_2^2) & \text{with probability } (1-p) \end{cases}, \quad \begin{matrix} 0 < p < 1 \\ p\sigma_1^2 + (1-p)\sigma_2^2 = 1 \end{matrix}$$

– The resulting kurtosis of ϵ_t is

$$\text{Kurt} = \frac{3[p\sigma_1^4 + (1-p)\sigma_2^4]}{[p\sigma_1^2 + (1-p)\sigma_2^2]^2} = 3[p\sigma_1^4 + (1-p)\sigma_2^4] \geq 3$$

– A third means of introducing fat tails into return distributions is to add a jump process to the data-generating process

– This can be accomplished by adding a Bernoulli random "jump" Z_t, taking on the values {0, 1} and satisfying $\Pr\{Z_t = 1\} = \zeta$, with independent random amplitude ξ, e.g., $\xi \sim N(m_J, \delta^2_J)$.

$$r_t = \mu_{t-1} + \sigma\epsilon_t + \xi Z_t$$

NON-AFFINE MODELS: ML ESTIMATION

- Bernoulli jump model has r_t jumping at most once btw. t − 1 and t
  - Given a parametric assumption about the distribution of ϵ_t conditional on the past history of returns, and assuming that a mixture or jump process is independent of ϵ_t , then all of these discrete-time volatility models can be estimated by MLE
  - Excluding jumps, because $r_t | I_{t-1} \sim N(\mu_{t-1}, \sigma^2_{t-1})$, the likelihood function of the data is

$$L_T(r_T, \ldots, r_1 | \vec{r}_0^{\,J}, \vec{\sigma}_0^{2J}; \theta) = f(r_T, \ldots, r_1 | \vec{r}_0^{\,J}, \vec{\sigma}_0^{2J}; \theta)$$

$$= f(r_T | I_{T-1}; \theta) f(r_{T-1} | I_{T-2}; \theta) \ldots f(r_1 | \vec{r}_0^{\,J}, \vec{\sigma}_0^{2J}; \theta)$$

$$f(r_t | I_{t-1}; \theta) = \frac{1}{\sqrt{2\pi\sigma^2_{t-1}}} e^{-\frac{1}{2}(r_t - \mu_{t-1})^2 / \sigma^2_{t-1}}$$

  - Alternatively, in the MixGARCH model, we have

NON-AFFINE MODELS: ML ESTIMATION

$$f(r_t|I_{t-1};\theta) = p \times \frac{1}{\sqrt{2\pi\sigma_{t-1}^2\sigma_1^2}} e^{-\frac{1}{2}(r_t-\mu_{t-1})^2/(\sigma_{t-1}^2\sigma_1^2)}$$

$$+ (1-p) \times \frac{1}{\sqrt{2\pi\sigma_{t-1}^2\sigma_2^2}} e^{-\frac{1}{2}(r_t-\mu_{t-1})^2/(\sigma_{t-1}^2\sigma_2^2)}$$

- Though easy to write down, the likelihood function of the Mixture GARCH model is globally unbounded, a well-known problem of mixture-of-normal models
  - Just set $\mu_0 = r_1$ and let σ_1 approach zero: the log-lik at t = 1 approaches infinity and, hence, so does the likelihood function
  - This is typically not a problem in practice, because numerical search routines find local optima, and one can search across local optima with bounded likelihood function values
- Kiefer (1978, ECMA) shows that there exists a consistent, asymptotically normal local optimum with usual MLE properties