Maximum Likelihood Estimation of Latent Affine Processes

David S. Bates

University of Iowa and NBER

Institute for Mathematics and Its Applications
Risk Management and Model Specifications Issues in Finance
April 16, 2004

Issue: Estimating continuous-time models with latent state variables on discrete-time data.

-Parameters of processes

-Realizations of latent variables

Examples: Stochastic volatility

Time-varying jump intensities

Existing approaches:

1) *Assume* latent variable realizations can be inferred from prices of derivatives

2) Tractable state-space models

-Gaussian structure + Kalman filtration

-Regime-switching models

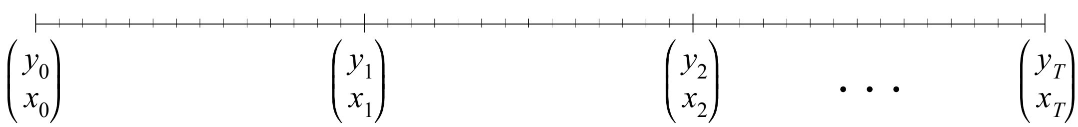
3) Simulation-based methods

- Simulated method of moments
- Gallant and Tauchen's EMM/SNP
- Bayesian MCMC methods (JPR '94, EJP '03)
- Particle filters

Gallant and Tauchen's SNP/EMM

- Derive moment conditions for (discrete-time) data from an auxiliary discrete-time model
  - -ARMA for first moments
  - -ARCH for second moments
  - -Hermite polynomials for fat-tailed behavior

- Simulate "continuous-time" data and latent variable realizations over a fine time grid for given parameters. Choose parameters that match the discrete-time moment conditions via GMM.



- ***Reprojection***: derive a filtration rule by a "kitchen sink" regression of simulated latent variable realizations on past simulated data realizations.

Concerns

- *Estimation:* How well do these simulation methods work with infrequent outliers?
- *Filtration:* rather cumbersome; not directly comparable to, e.g., GARCH methods

My approach:

- Analogous to the filtration approaches of Gaussian & regime-switching models
  - -Estimates of latent state variable(s) updated recursively conditional on observed data & parameter values.

- ***Major innovation:*** updating occurs in the transform space of *characteristic functions*, rather than in probability densities.

- ***Major limitations*:**

  Requires for "semi-affine" processes

  Curse of dimensionality

  But: These are interesting models, and are convenient for pricing bonds and options.

  -affine stochastic volatility (SV)

  -SV + jumps in asset prices (SVJ)

  -SV + jumps in asset prices *correlated* with jumps in volatility

  -Stochastic jump intensities

  -Multifactor models

  -Time-changed Lévy processes

  Some non-affine models become affine after data transformation

  -discrete-time log volatility of HRS

Outline

- Basic algorithm

- Diagnostics on simulated data
  - -filtration efficiency
  - -parameter estimation efficiency

- Estimates on daily stock index returns, 1953-96, of affine stochastic volatility models without and with jumps

- Option pricing applications

1. Basic algorithm

Fourier Transforms

Let $$F(i\Phi, i\psi) \equiv \iint e^{i\Phi y + i\psi x} p(y, x)\, dx\, dy$$

be the joint characteristic function of $(\tilde{x}, \tilde{y})$.
(Analytic for affine models)

Inverse Fourier transforms

Marginal densities: $$p(y) = \frac{1}{2\pi} \int F(i\Phi, 0)\, e^{-i\Phi y}\, d\Phi$$

$$p(x) = \frac{1}{2\pi} \int F(0, i\psi)\, e^{-i\psi x}\, d\psi$$

Joint density functions:

$$p(y, x) = \frac{1}{(2\pi)^2} \iint F(i\Phi, i\psi)\, e^{i\Phi y - i\psi x}\, d\Phi\, d\psi .$$

Proposition 1: *The characteristic function of x* ***conditional*** *upon observing y is a partial inversion of F:*

$$G_{x\,|\,y}(i\psi \mid y) = \frac{\frac{1}{2\pi} \int_{-\infty}^{\infty} F(i\Phi, i\psi)\, e^{-i\Phi y}\, d\Phi}{p(y)} \qquad \textbf{(3)}$$

(Bartlett, 1938).

***Proof*:** By Bayes' law, the conditional characteristic function $G_{x\,|\,y}$ can be written as

$$G_{x\,|\,y}(i\psi;\ y) \;=\; \int e^{i\psi x}\, p(x\,|\,y)\, dx$$

$$\;=\; \frac{1}{p(y)} \int e^{i\psi x}\, p(y, x)\, dx\,.$$

$F(i\Phi, \bullet)$ is therefore the (Φ, y) Fourier transform of $G_{x\,|\,y}(\bullet, y)\, p(y)$:

$$F(i\Phi, i\psi) \;=\; \int e^{i\Phi y}\, [\, G_{x\,|\,y}(i\psi, y)\, p(y)]\, dy$$

$$\;=\; \int\!\!\int e^{i\Phi y \,+\, i\psi x}\, p(y, x)\, dx\, dy\,.$$

Consequently, $G_{x\,|\,y}(i\psi, y)\, p(y)$ is the *inverse* (Φ, y) Fourier transform of $F_{x,\,y}(i\psi, i\Phi)$, yielding Proposition 1. ■

Semi-affine models

Assumption:
Thejoint CF for the distribution of future realizations of the data $\tilde{y}_{t+1}$ and the latent variable $\tilde{x}_{t+1}$ has an analytic, exponentially affine solution conditional upon knowing x_t:

$$\begin{aligned} F(i\Phi,\ i\psi \mid y_t, x_t) &\equiv E[e^{i\Phi y_{t+1} + i\psi x_{t+1}} \mid y_t, x_t] \\ &= \exp[C(\Delta t, i\Phi, i\psi;\ y_t) + D(\Delta t, \Phi, \psi)\, x_t] \end{aligned}$$

Given this and Proposition 1, the econometrician's joint CF $F(i\Phi,\ i\psi \mid \boldsymbol{Y}_t)$ conditional upon observing only past data $\boldsymbol{Y}_t = \{y_1, \ldots, y_t\}$ can be updated recursively over time.

Procedure:
Let

$$G_{t|t}(i\psi) \equiv E\left[e^{i\psi x_t} \mid \boldsymbol{Y}_t\right]$$

summarize current knowledge regarding the latent variable x_t. Initial value:

$$\begin{aligned} G_{0|0}(i\psi) &= E[\exp(i\psi x_0)] \\ &= \exp[C(\Delta t = \infty, 0, i\psi)] \end{aligned}$$

Using Proposition 1 and the affine structure, $G_{t|t}$ can be updated recursively.

Table 1: Fourier inversion recursion to computing likelihood functions

Cond. CF: $F(i\Phi,\ i\psi \mid y_t, x_t) \equiv E[e^{i\Phi y_{t+1} + i\psi x_{t+1}} \mid y_t, x_t] = \exp[C(i\Phi, i\psi) + D(i\Phi, i\psi)\, x_t]$

Densities	Associated characteristic functions
Conditional density $p(x_t \mid t)$	$G_{t\mid t}(i\psi) \equiv E_t\left(e^{i\psi x_t}\right)$
Joint density of data and latent variable: $p(y_{t+1}, x_{t+1} \mid \boldsymbol{Y}_t) = \int p(y_{t+1}, x_{t+1} \mid y_t, x_t)\, p(x_t \mid \boldsymbol{Y}_t)\, dx$	$F(i\Phi, i\psi \mid \boldsymbol{Y}_t) = E_t\left[E\left(e^{i\Phi y_{t+1} + i\psi x_{t+1}} \mid x_t\right)\right] = E_t\left[e^{C(i\Phi, i\psi) + D(i\Phi, i\psi)\, x_t}\right] = e^{C(i\Phi, i\psi)}\, G_{t\mid t}[D(i\Phi, i\psi)]$
Evaluation of likelihood function: $p(y_{t+1} \mid \boldsymbol{Y}_t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(i\Phi, 0 \mid \boldsymbol{Y}_t)\, e^{-i\Phi y_{t+1}}\, d\Phi$	
Updated conditional density of x_{t+1}: $p(x_{t+1} \mid \boldsymbol{Y}_{t+1}) = \frac{p(y_{t+1}, x_{t+1} \mid \boldsymbol{Y}_t)}{p(y_{t+1} \mid \boldsymbol{Y}_t)}$	$G_{t+1\mid t+1}(i\psi) = \frac{\frac{1}{2\pi}\int_{-\infty}^{\infty} F(i\Phi, i\psi \mid \boldsymbol{Y}_t)\, e^{-i\Phi y_{t+1}}\, d\Phi}{p(y_{t+1} \mid \boldsymbol{Y}_t)}$

Noncentral moments of x_{t+1}: $E[x_{t+1}^n \mid t+1] = \left.\frac{\partial^n G_{t+1\mid t+1}(\psi)}{\partial \psi^n}\right|_{\psi = 0}$

Approximating functions $\hat{G}_{t|t}(i\psi)$

$G_{t|t}(i\psi)$ summarizes all that is known about x_t at time *t*.

Needs to be temporarily *stored* , in a fashion accessible for arbitrary (complex-valued) inputs.

Approximating function $\hat{G}_{t|t}(i\psi)$

Numerical Recipes, Judd:

-Splines, Chebychev polynomials

-***But:*** resulting CF's may not yield valid densities

Approach here: ***approximate maximum likelihood*** **(AML)**

- Use $\hat{G}_{t|t}(i\psi)$ from known density functions to impose shape constraints

  -gamma: $\ln G_{t|t}(i\psi) \approx -\nu_t \ln(1 - i\kappa_t\psi)$

  -normal: $\ln G_{t|t}(i\psi) \approx i\psi\,\hat{x}_{t|t} + \tfrac{1}{2}(i\psi)^2 P_{t|t}$

  depending on support of latent x_t.

- Update moments and associated parameters from recursion

$$\hat{x}_{t+1|t+1} = G'_{t+1|t+1}(0)$$

$$P_{t+1|t+1} = G''_{t+1|t+1}(0) - \hat{x}^2_{t+1|t+1}$$

Analogous to Kalman filtration's updating of moments; *but*

- Approach is using the optimal *nonlinear* updating rule (conditional upon the prior being approximately correct); and
- the resulting filtration is *numerically stable.*

Test of approach on affine simulated data

Discrete-time stochastic volatility process

$$z_{t+1} = \sqrt{V_t}\,\tilde{\varepsilon}_{t+1}$$

$$\ln V_{t+1} = \omega + \varphi \ln V_t + \sigma_v \tilde{\eta}_{t+1}$$

(affine under the transformation $y_{t+1} = \ln z_{t+1}^2$)

Continuous-time jump-diffusion

$$dS/S = (\mu_0 + \mu_1 V - \lambda_t \bar{k})\,dt + \sqrt{V}\,dW + (e^{\gamma} - 1)\,dN$$

$$dV = (\alpha - \beta V)\,dt + \sigma\sqrt{V}\,dW_V$$

$$Corr[dW, dW_V] = \rho$$

$$Prob[dN = 1] = (\lambda_0 + \lambda_1 V)\,dt, \quad \gamma \sim N[\bar{\gamma}, \delta^2]$$

Filtration

Discrete-time log variance, Gaussian prior

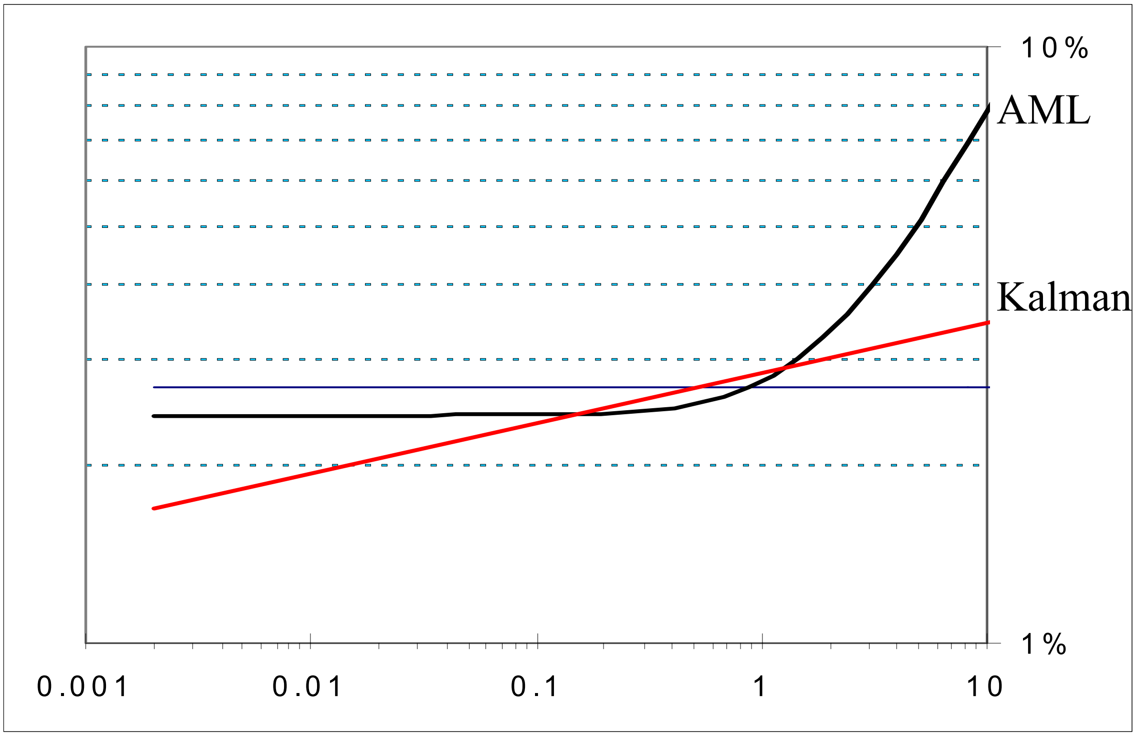


Figure B.1. Weekly volatility estimates $E_{t+1}\sqrt{V_{t+1}}$ conditional upon seeing an absolute return of size $|z_{t+1}|/E_t\sqrt{V_t}$. Log scales on both axes.

Filtration efficiency for weekly returns

(R^2 from regressing true volatility on filtered estimates)

Kalman: 27%

AML: 37-38%

Volatility filtration for continuous-time SV/jump models

-Approximate prior: gamma

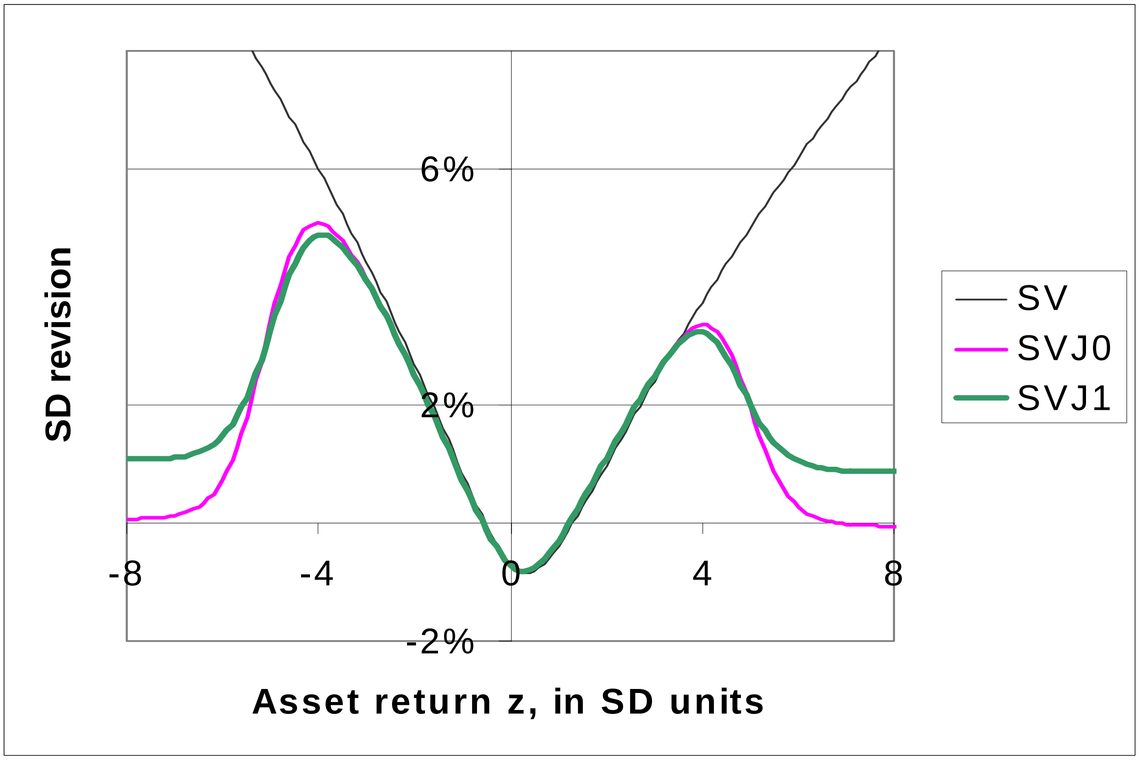


Figure 1: News impact curves for continuous-time models. The graph shows the revision in assessed conditional standard deviation, $(E_{t+1} - E_t)\sqrt{V_{t+1}}$, conditional upon observing an standardized asset return of magnitude $z = y_{t+1} / \sqrt{V_{t|t}\Delta t}$.

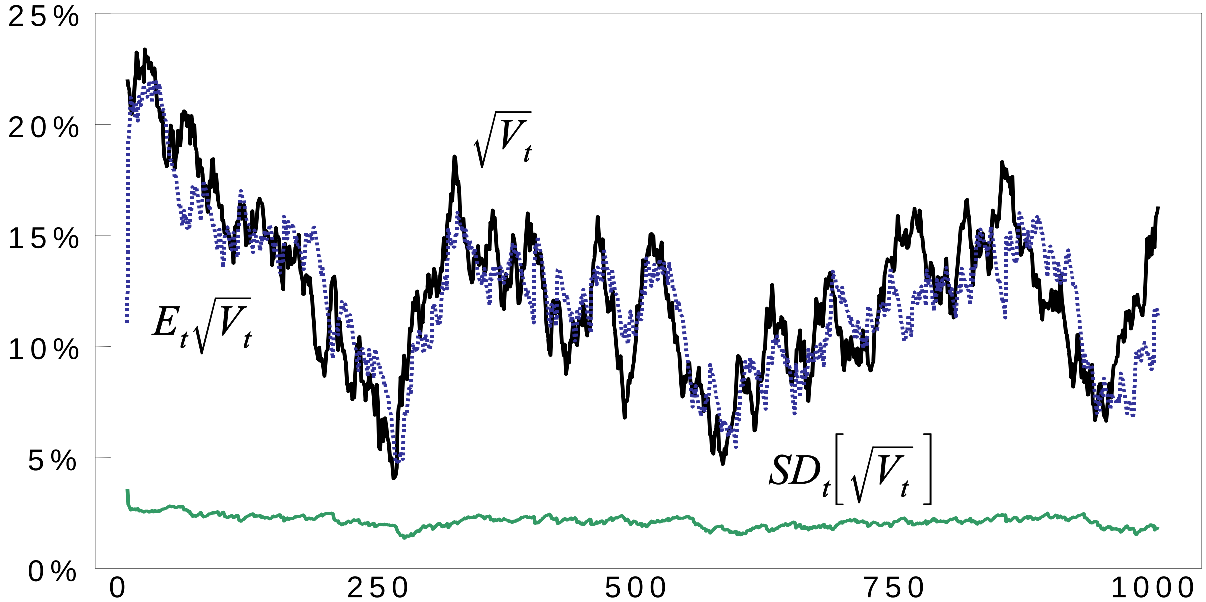


Table 5: Volatility and variance filtration efficiency

Filtration method	R^2 on SV data: volatility	R^2 on SV data: variance	Filtration method	R^2 on SVJ1 data: volatility	R^2 on SVJ1 data: variance
In-sample fits					
GARCH	.598	.553	*t*-GARCH	-.008	-2.222
EGARCH	.674	.649	*t*-EGARCH	.356	-2.184
HGARCH	.678	.649	*t*-HGARCH	.585	.453
SV ($\boldsymbol{\theta}$)	**.703**	**.690**	SVJ1 ($\boldsymbol{\theta}$)	**.723**	**.700**
SV ($\hat{\boldsymbol{\theta}}$)	.704	.692	SVJ1 ($\hat{\boldsymbol{\theta}}$)	.726	.702
SV($\boldsymbol{\theta}$) + reprojection		.695	SVJ1($\boldsymbol{\theta}$) + reprojection		.705
Out-of-sample fits					
HGARCH	.678	.650	*t*-HGARCH	.603	.430
SV ($\boldsymbol{\theta}$)	**.699**	**.687**	SVJ1 ($\boldsymbol{\theta}$)	**.747**	**.727**
SV($\boldsymbol{\theta}$) + reprojection		.689	SVJ1($\boldsymbol{\theta}$) + reprojection		.727

Parameter estimation efficiency

	T = 500 weeks			T = 2000 weeks		
True values:	ω -.736	φ .90	σ_v .363	ω -.736	φ .90	σ_v .363
Root mean squared errors						
QML	1.60	.22	.27	.46	.06	.11
GMM	.59*	.08*	.17*	.31	.04	.12
EMM	.60	.08	.20	.224	.030	.049
AML	**.42**	**.06**	**.08**	**.173**	**.023**	**.043**
MCMC	.34	.05	.07	.15	.02	.034
ML	.43	.05	.08	NA	NA	NA
MCL	.27	.04	.08	NA	NA	NA

ML: Fridman and Harris (1996)
MCL: Sandmann and Hoopman (1996)

Parameter estimation for continuous-time SV/jump model
-About the same RMSE as MCMC
-Less bias

In sum:

- AML more efficient than EMM, and about as efficient as MCMC.
- AML directly provides filtered estimates of latent variable realizations.

Estimates from stock return data, 1953-96

$$dS/S = (\mu_0 + \mu_1 V - \lambda_t \bar{k})\, dt + \sqrt{V}\, dW + (e^{\gamma} - 1)\, dN$$

$$dV = (\alpha - \beta V)\, dt + \sigma \sqrt{V} dW_V$$

$$Corr[dW, dW_V] = \rho$$

$$Prob[dN = 1] = (\lambda_0 + \lambda_1 V)\, dt, \quad \gamma \sim N[\bar{\gamma}, \delta^2]$$

$$\bar{k} = \exp[\bar{\gamma} + \tfrac{1}{2}\delta^2] - 1$$

Issues:

- What's the magnitude of jump risk?

- Do jump intensities vary with the level of volatility?
  - Bates (2000): better matches option prices.
  - Mixed time series results from Bates & Craine (1999) Eraker *et al* (2000), and ABL (2002).

- Does volatility jump?
  (Bates, 2000: ISD's from option prices jump)

Data

Andersen, Benzoni, & Lund (*JF,* 2002):

-11,076 daily log-differenced S&P 500 prices over 1953 - 96.

-Prefiltered to remove an MA(1) component

-3 major negative outliers:

- October 19, 1987: -22.22%
- September 26, 1955 -6.99%
- October 13, 1989 -6.31%

- October 20, 1987: +9.02%
- October 21, 1987: +7.25%

<u>Estimates using EMM/SNP</u>

ABL: Daily S&P 500 returns, 1953 - 96 (11,076 obs)

CGGT: Daily DJIA returns, 1953 - July 16, '99 (11,717 obs)

Table 7: Model estimates, and comparison with EMM-based results

Model:

$$dS/S = (\mu_0 + \mu_1 V - \lambda_t \bar{k})\,dt + \sqrt{V}\left(\rho\, dW_1 + \sqrt{1-\rho^2}\, dW_2\right) + (e^{\gamma_s} - 1)\,dN$$

$$dV = (\alpha - \beta V)\,dt + \sigma\sqrt{V}\,dW_1$$

$$Prob[dN = 1] = (\lambda_0 + \lambda_1 V)\,dt, \quad \gamma_s \sim N[\bar{\gamma}, \delta^2]$$

CGGT: Chernov *et al* (2003) EMM-based estimates on daily DJIA returns, 1953 - July 16, 1999 (11,717 obs.).
ABL: Andersen et al (2002) EMM-based estimates on daily S&P 500 returns, 1953 -1996 (11,076 obs.).
AML: Approximate maximum likelihood estimates of this paper, using the ABL data set.

All parameters in annualized units except the variance shock half-life $HL = 12\ln 2/\beta$, which is in months.

	SV			SVJ0, $\lambda_1 = 0$			SVJ1, $\lambda_1 \neq 0$		
	CGGT	ABL	**AML**	CGGT	ABL	**AML**	ABL	**AML**	**AML**
μ_0		.051	**.026**		.037	**.028**	.037	**.040**	**.040**
		(.032)	**(.025)**		(.045)	**(.027)**	(.095)	**(.025)**	**(.025)**
μ_1		2.58	**3.70**		4.02	**3.89**	4.03	**3.09**	**3.09**
		(2.82)	**(1.98)**		(3.89)	**(2.19)**	(5.77)	**(2.16)**	**(2.16)**
α	1.283	.051	**.093**	.044	.047	**.063**	.047	**.061**	**.061**
		(.010)	**(.011)**		(.013)	**(.009)**	(.017)	**(.008)**	**(.008)**
β	137.87	3.93	**5.94**	2.79	3.70	**4.38**	3.70	**4.25**	**4.25**
	(.17)	(.81)	**(.81)**	(.54)	(1.08)	**(.70)**	(1.71)	**(.59)**	**(.59)**
σ	1.024	.197	**.315**	.207	.184	**.244**	.184	**.237**	**.237**
	(.030)	(.018)	**(.018)**	(.02)	(.019)	**(.016)**	(.019)	**(.015)**	**(.015)**
ρ	-.199	-.597	**-.579**	-.483	-.620	**-.612**	-.620	**-.611**	**-.611**
	(.000)	(.045)	**(.031)**	(.10)	(.067)	**(.031)**	(.086)	**(.031)**	**(.031)**
$\sqrt{\alpha/\beta}$	.096	.114	**.125**	.125	.113	**.120**	.113	**.119**	**.119**
			(.004)			**(.004)**		**(.004)**	**(.004)**
HL	0.06	2.12	**1.40**	2.98	2.25	**1.90**	2.25	**1.96**	**1.96**
	(.00)	(.44)	**(.19)**	(.58)	(.66)	**(.31)**	(1.04)	**(.27)**	**(.27)**
λ_0				1.70	5.09	**.744**	5.09		**.000**
					(.43)	**(.217)**	(7.18)		**(.000)**
λ_1							.70	**93.4**	**93.4**
							(488.0)	**(33.4)**	**(33.4)**
$\bar{\gamma}$				-.030		**-.010**		**-.002**	**-.002**
				(.002)		**(.010)**		**(.006)**	**(.006)**
δ				.008	.012	**.052**	.012	**.039**	**.039**
				(.001)	(.001)	**(.009)**	(.001)	**(.008)**	**(.008)**
ln L		39,192.45[a]	**39,233.87**		39,238.03[a]	**39,294.79**	39,238.03[a]	**39,309.51**	**39,309.51**

[a]ABL log likelihoods were evaluated at the ABL parameter estimates using the AML methodology

Summary of Results

- Strong evidence that jumps are more likely when volatility is high
- EMM/SNP estimates diverge significantly from AML estimates.
- EMM/SNP estimates' sensitivity to infrequent outliers is strongly affected by the specification of the auxiliary model – especially for SV process.

- ML jump risk estimates: Less frequent, more substantial jumps
- Jumps drawn from a mixture of normals capture the '87 crash better.

$$Prob[dN_1 = 1] = 131.1\, V_t, \quad \ln(1 + k_1) \sim N[\ .001,\ (.029)^2]$$

$$Prob[dN_2 = 1] = 2.4\, V_t, \quad \ln(1 + k_2) \sim N[-.222,\ (.007)^2].$$

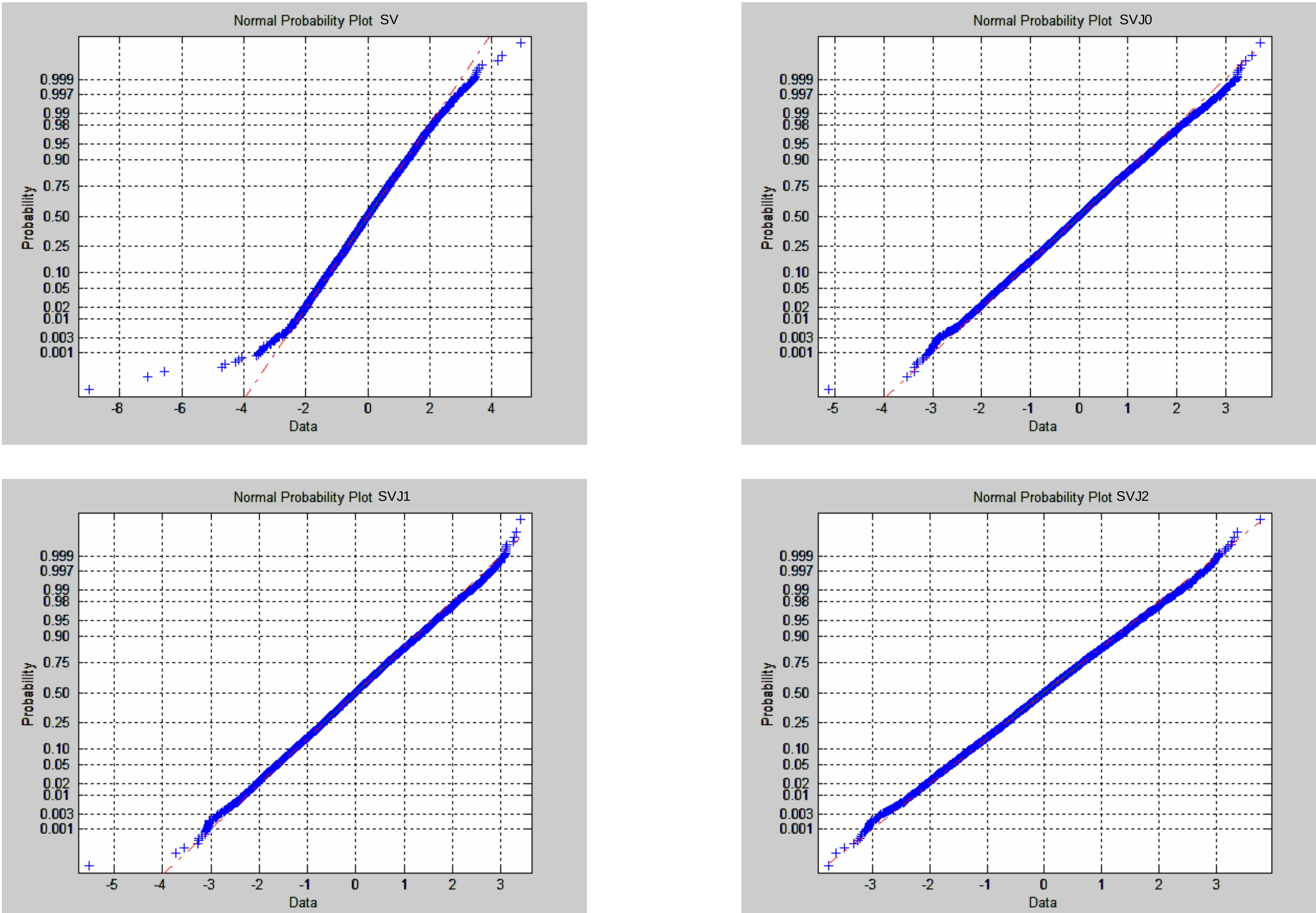


Figure 3: Normal probability plots for the normalized returns $y^*_{t+1} \equiv N^{-1}[CDF(y_{t+1}\,|\,Y_t, \hat{\theta})]$, for different models. The diagonal line gives the theoretical quantiles conditional upon correct specification; + gives the empirical quantiles.

Option pricing

Option pricing conditional upon knowing V_t:

$$c(S_t, V_t, T; X) = S_t e^{-d_t T} - e^{-rT} X \left[\frac{1}{2} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{C^*(i\Phi, 0) + D^*(i\Phi, 0) V_t - i\Phi \ln(X/S_t)]}}{i\Phi(1 - i\Phi)} d\Phi \right]$$

Econometrician's valuation:

$$\begin{aligned} c(S_t, T; X \mid \boldsymbol{Y}_t) &= E[c(S_t, V_t, T; X) \mid \boldsymbol{Y}_t] \\ &= S_t e^{-d_t T} - e^{-rT} X \left[\frac{1}{2} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{C^*(i\Phi, 0)} G_{t|t}[D^*(i\Phi, 0)] e^{-i\Phi \ln(X/S_t)}}{i\Phi(1 - i\Phi)} d\Phi \right] \end{aligned}$$

General affine pricing kernel

$$d \ln M = \mu_m dt - R\, d \ln S - R_V dV - R_J \gamma_s dN$$

Here, use $R_V = R_J = 0$,
and R estimated from time series model

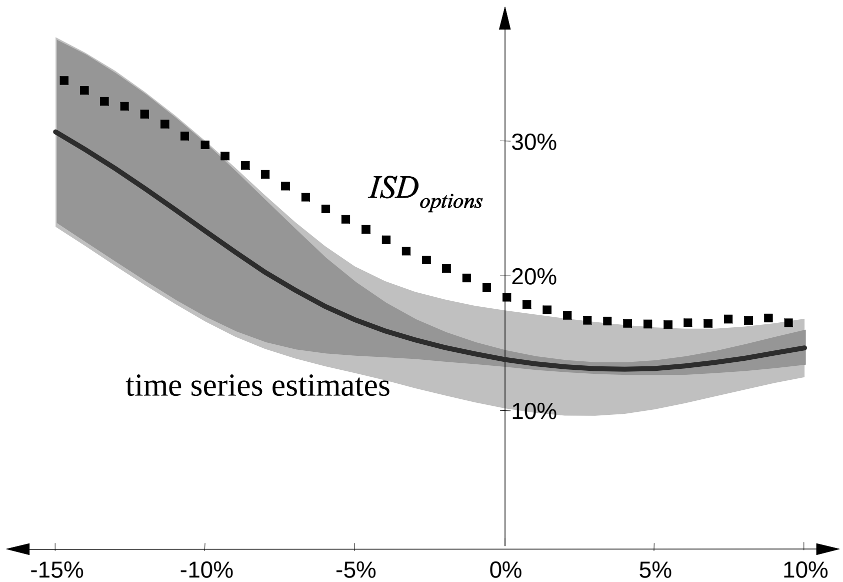


Figure 6. Observed and estimated ISD's for 17-day Jan. '97 S&P 500 futures options on December 31, 1996. The dark grey area is the 95% confidence interval given only parameter estimation uncertainty. The light grey area is the 95% confidence interval given both parameter and state uncertainty.

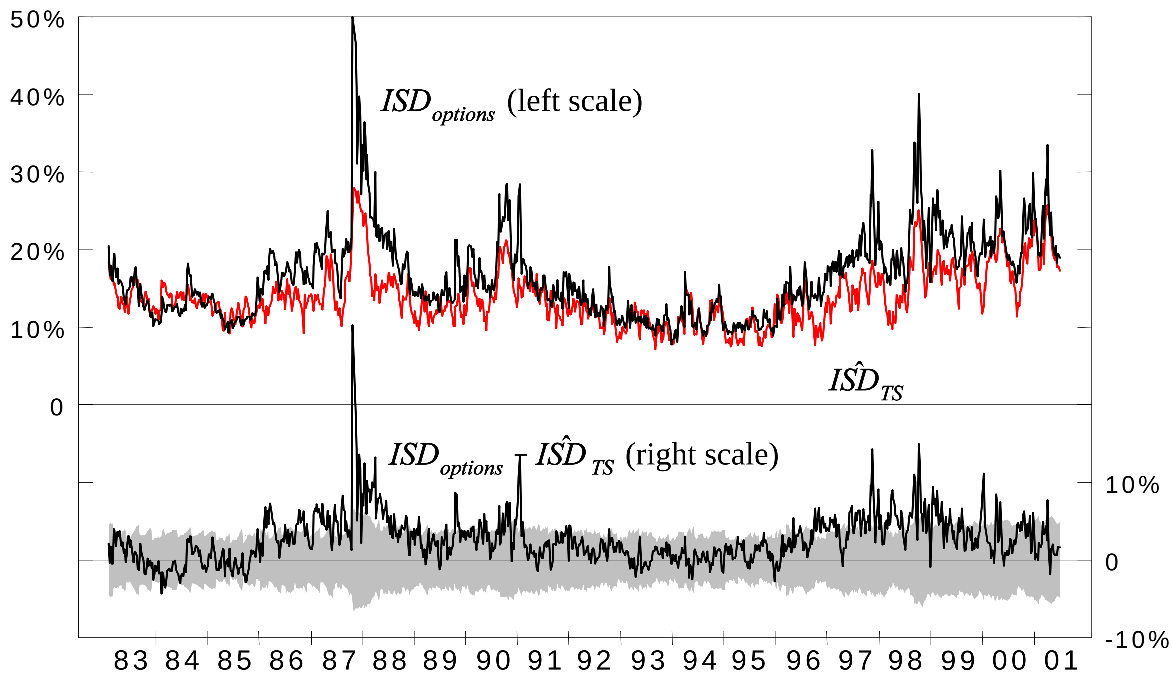


Figure 7. Observed and estimated ISD's for at-the-money S&P 500 futures options. Estimated ISD's are based on SVJ2 parameter estimates from 1953-96, and on filtered variance estimates. Gray area is the 95% confidence interval for $ISD_t - I\hat{S}D_t$, given both parameter and state uncertainty.

Potential extensions

Bayesian

likelihood + prior => posterior

Recent extended-data stochastic volatility approaches also fit within the affine structure

-"Realized" (or integrated) variance

-Implicit variances from option prices

-Range-based variance

Other applications

Any low-dimensional affine state space system; e.g., autoregressive conditional duration (ACD)