Department of Economics

Continuous Time Regime Switching Models and Applications in Estimating Processes with Stochastic Volatility and Jumps

Kyriakos Chourdakis

Working Paper No. 464 November 2002 ISSN 1473-0278

Continuous Time Regime Switching Models and Applications in Estimating Processes with Stochastic Volatility and Jumps

Kyriakos Chourdakis*

October 2002

Abstract

A regime switching model in continuous time is introduced where a variety of jumps are allowed in addition to the diffusive component. The characteristic function of the process is derived in closed form, and is subsequently employed to create the likelihood function. In addition, standard results of the option pricing literature can be employed in order to compute derivative prices. To this end, the relationship between the physical and the risk adjusted probability measure is explored. The generic relationship between Markov chains and [jump] diffusions is also investigated, and it is shown that virtually any stochastic volatility model model can be approximated arbitrarily well by a carefully chosen continuous time Markov chain. Therefore, the approach presented here can be utilized in order to estimate, filter and carry out option pricing for such continuous state-space models, without the need for simulation based approximations. An empirical example illustrates these contributions of the paper, estimating a stochastic volatility jump diffusion model.

Keywords: Continuous time regime switching, stochastic volatility jump diffusion, option pricing, filtering
JEL Classification: G10, G13, C22

* k.chourdakis@qmul.ac.uk. ✉ Room W306 Department of Economics, University of London, Queen Mary, London E1 4NS. ☎ ++44 +207 882 5086. http:\\www.qmul.ac.uk\ ~te9001, where software that utilizes the approaches discussed in this paper for MATHEMATICA 4.1 can be downloaded from. Comments most welcome.

REGIME SWITCHING MODELS have already enjoyed much success in interpreting the behavior of a number of economic and financial series in a concise, yet parsimonious way. After the seminal work of Hamilton (1989), a vast number of researchers have utilized the regime switching approach to model virtually every time series encountered in the economic literature. The appeal of the filter presented in Hamilton (1989) is based on two important characteristics: (i) The filter can be viewed as a discrete state space/discrete time version of the Kalman–Bucy filter, and its implementation and calculation of the likelihood function over a discrete sample is straightforward, and (ii) as a byproduct of the maximum likelihood estimation procedure one obtains the filtered probability distribution of the unobserved process, conditional on the information which prevails at the time.

This paper attempts to generalize such processes for the continuous time case. A Lévy process is used as the instrument that models the evolution of a phenomenon, where the parameters of this Lévy process are allowed to depend on the state of an unobserved Markov chain that lives in continuous time. The choice of a Lévy process is based on its ability to encompass specifications with both continuous and discontinuous sample paths, while maintaining a sufficient level of mathematical tractability. The results presented in this paper rely heavily on the analytical form of the characteristic function of a Lévy process, given by the famous Lévy–Khinchine formula and its variants [for details see Bertoin (1996)]. The first contribution of this paper is the the characteristic function of a state dependent Lévy process, which is derived in closed form.

Having obtained the characteristic function of the process under consideration opens a number of ways of exploring its properties and construct estimation procedures. Among others, one convenient approach which is analogous to the one in Hamilton (1989) is presented. Using such a methodology allows the researcher to retrieve the conditional distribution of the unobserved Markov chain process, while the procedure remains tractable and easily implementable. A number of ways that can be employed in order to speed up computations, such as Hermite,

logspline and Pearson's IV density approximations are also discussed. In addition, the characteristic function can be used in order to price derivative contracts, utilizing the results presented in Bakshi and Madan (2000).

The most popular specifications employed for modelling financial time series and for derivative pricing include the ones where jumps and stochastic volatilities are present. Such model have enjoyed much success, mainly due to their ability to replicate the stylized facts of asset returns, which are summarized in the Black–Scholes implied volatility smile [see Ghysels, Harvey, and Renault (1996) for details on the stylized facts and the smile]. Although stochastic volatility models have been attractive theoretically, there have been a number of difficulties concerning the effective estimation of their parameters, resulting from the fact that the transition density is not readily available in closed form. Simulation methods[1] are normally employed, with the EMM approach of Gallant and Tauchen (1997) most widely used.[2]

The Markov chain approach examined in this paper can offer an alternative methodology in estimating processes with latent diffusions. Intuitively, a Markov chain has the generic structure which is shared with the majority of stochastic volatility models, or other specifications with unobserved latent factors. The second contribution of the paper is to examine the ways that this relation can be explored in order to estimate, filter and carry out option pricing for such continuous state-space models. It is shown that virtually any stochastic volatility model model can be approximated arbitrarily well by a carefully chosen continuous time Markov chain. Thus, estimation, filtering and option pricing for stochastic volatil-

[1]For specific cases approximate linear filters have been utilized, in the spirit of Harvey, Ruiz, and Shephard (1994). ARCH filters have also served as approximations of the unobserved volatility process, based on the results of Nelson (1990) and Nelson and Foster (1994). The Markovian structure of the volatility process has been exploited in the Bayesian approach of Jacquier, Polson, and Rossi (1995) which is simulation based.

[2]The EMM approach uses simulations based on the seminonparametric density expansion discussed in Gallant and Tauchen (1997). Recent examples of this methodology can be found in Chernov, Gallant, Ghysels, and Tauchen (1999) and Chernov, Gallant, Ghysels, and Tauchen (2002).

ity models can be carried out using the simple and numerically convenient Markov chain approximation.

The researcher can apply the maximum likelihood methodology discussed here in order to estimate the parameters of this approximating process, taking advantage of its full efficiency. Kushner (1990) and Dupuis and Kushner (2001) give excellent overviews of such approximation procedures in a stochastic control context. This paper discusses the ways such procedures can be applied in order to estimate models with stochastic volatilities and jumps, and a number of extensions of the standard model are also overviewed, including volatility that depends on a vector of factors, jumps in the volatility process and feedback effects from the asset price process on the volatility diffusion. It is important to note that models permitted in the context of this research include ones where the parameters are affine [in the spirit of Duffie, Pan, and Singleton (1999) and Pan (2002)] but models with nonaffine parameters are also allowed.

As noted before, since the characteristic function is derived in closed form, derivative prices can be computed which are exact up to a numerical integration error. Bakshi and Madan (2000) show that contracts with trigonometric payoffs span the contingent claims space, and give a general formula that expresses the price of a European option as a function of the Fourier transform of the risk neutral density. Following the comments of the above paragraph, one can implement the results of this paper in order to compute approximate derivative prices for stochastic volatility models with jumps. Although a direct application would be the computation of prices that do not admit explicit solutions [such as the logarithmic model of Scott (1987)], affine models could also benefit from the closed form approximation presented here rather than the numerical approach of Duffie, Pan, and Singleton (1999).

A sample that spans ten years from April 1987 to December 1997 of the SP500 index is used to apply the results of the paper. A family of models exhibiting stochastic volatility model with affine jumps is estimated. The choice of the SP500

is based on two facts: (i) This index has been the subject of a vast empirical research, allowing one to compare the results of this estimation procedure with the ones reported in the literature, and (ii) the index serves as the underlying asset for the SPX options, a family of very liquid contracts. In that fashion, the results can also be compared to the ones obtained from filtering option prices alone. The choice of the sample size is based on the comments of Bates (1997), in particular on the fact that the behavior of derivative prices has changed dramatically following the crash of 87. Specifications with and without correlations and jumps are used, in order to explore the effects of these alternatives on the parameter estimates.

It is found that the estimated parameters of this paper are in line with the ones reported in the literature. The jumps are rare events [average 1.7 jumps per year] and they are expected to be negative with sizes that are proportional to the volatility [average jump size -0.6%]. Therefore, highly volatile periods are found to exhibit not only more frequent jumps, but more severe ones as well. The correlation coefficient which is responsible for the skewness encountered in the data is estimated to be around -0.60, a value verified by a number of studies [for example Dumas, Fleming, and Whaley (1998) based on option prices and Andersen, Benzoni, and Lund (1998) based on time series data].

A byproduct of the estimation algorithm presented here is the filtered distribution of the latent state. Based on this distribution a number of observations are made: Highly volatile market periods seem to be accompanied by higher certainty of the volatility level. In contrast, when the market is quiet the agents seem to have a higher degree of uncertainty on the exact level. Such observations naturally lead to the informal discussion about the impact of this volatility uncertainty on option prices.

The plan of the paper is the following: Section 2 introduces the continuous time regime switching model. The main technical results can be found in section 3, which derives the characteristic function and computes the conditional and un-

conditional moments of the state dependent process. Section 3 extends the filter of Hamilton (1989) for the continuous time case, while section 4 discusses how derivative contracts are priced in that setting. Section 5 attempts to bridge the gap between the model introduced here and the popular stochastic volatility jump diffusions. It is shown how the approximate chain is constructed, and a number of useful extensions are briefly discussed. The estimation procedure and the corresponding results are presented in section 6. Section 7 concludes and identifies a number of interesting issues for further research.

1 A class of state dependent processes

The underlying Markov chain. Consider a continuous–time Markov chain $\mathbf{x}(t)$, with rate matrix $\mathbf{Q}$ that lives in the orthonormal basis $\mathcal{B}$ of $\mathbb{R}^N$. The infinitecimal transition probabilities are given by

$$\mathsf{P}\left\{\mathbf{x}(t) = \mathbf{e}_j | \mathbf{x}(t-) = \mathbf{e}_i\right\} = \begin{cases} q_{ji}\mathrm{d}t + o(\mathrm{d}t) & , \text{if } \mathbf{e}_i \neq \mathbf{e}_j \\ 1 + q_{ii}\mathrm{d}t + o(\mathrm{d}t) & , \text{if } \mathbf{e}_i = \mathbf{e}_j \end{cases}$$

with the convention that $q_{ii} = -\sum_{j,j\neq i} q_{ji}$. The reason of choosing the unit vectors of being the state space of the Markov chain will be illustrated later, when the parameters of the various processes are specified.

The stochastic process. The paper deals with stochastic differential equations that can be disentangled in the form (1) together with assumptions (A1–A3) below, which has parameters that depend explicitly on the behavior of $\mathbf{x}(t)$, and discusses its properties based on closed form representations of the characteristic function of $S(t)$.

$$S(t) = s + Y(t) + J(t) + Z(t), \tag{1}$$

In (1) the three [indepented] stochastic integrals are meant to be translated in

the following way:

A1 The process $Y(t)$ is a state dependent Brownian motion, with instantaneous drift $\mu(\mathbf{x}(t-))$ and volatility $\sigma(\mathbf{x}(t-))$, and can therefore be represented as

$$Y(t) = \int_0^t \mu(\mathbf{x}(u-))\,\mathrm{d}u + \int_0^t \sigma(\mathbf{x}(u-))\,\mathrm{d}W(u)$$

with $W(t)$ a standard Wiener process. The drift and diffusion functions can take the form $\mu(\mathbf{x}(t)) = \mu^{\mathsf{T}}\mathbf{x}(t)$, and $\sigma(\mathbf{x}(t)) = \sigma^{\mathsf{T}}\mathbf{x}(t)$, with $\mu, \sigma \in \mathbb{R}^N$. For definiteness and without loss of generality, one can assume that the elements of σ are arranged in an increasing order.

A2 The process $J(t)$ is a compensated jump process, which jumps with intensity $\lambda(\mathbf{x}(t-))$ and exhibits a jump size of random magnitude $\nu_J(\mathbf{x}(t-))$ with associated measure $\nu_J(\mathbf{x}(t-), \cdot)$ on $\mathcal{B} \otimes A$, $A \subseteq \mathbb{R}$. It can be represented in terms of a random Poisson measure Π_J on $\mathcal{B} \otimes A$

$$J(t) = \int_0^t \int \Pi_J(\mathbf{x}(u-) \times \mathrm{d}\alpha)\mathrm{d}u - \lambda(\mathbf{x}(u-))\mathsf{E}\{\nu_J(\mathbf{x}(u-))\}\mathrm{d}u$$

where the innermost integral extends over the set A.

A3 The process $Z(t)$ jumps *only* when the underlying chain switches states, and exhibits a jump size of random magnitude $\nu_Z(\mathrm{d}\mathbf{x}(t))$ with associated measure $\nu_Z(\mathrm{d}\mathbf{x}(t) \times \cdot)$ on $\bar{\mathcal{B}} \otimes A$, $A \subseteq \mathbb{R}$, with $\bar{\mathcal{B}} = \{\beta \mid \exists\, \mathbf{e}, \bar{\mathbf{e}} \in \mathcal{B} : \beta = \mathbf{e} - \bar{\mathbf{e}}\}$ the set of all possible chain changes. This process can also be represented in a stochastic integral form

$$Z(t) = \int_0^t \int \Pi_Z(\Delta\mathbf{x}(u) \times \mathrm{d}\alpha)\mathrm{d}u - \sum_{\beta \in \bar{\mathcal{B}}} \mathsf{P}\{\Delta\mathbf{x}(u) = \beta | \mathbf{x}(u-)\}\mathsf{E}\{\nu_Z(\beta)\}\mathrm{d}u$$

where the Poisson measure Π_Z applies on $\bar{\mathcal{B}} \otimes A$, and the innermost integral again extends over A. The convention is now that ν_Z [and as an extension

$\Pi_Z]$ measures 0 a.s. when $\Delta\mathbf{x}(t) = \mathbf{0}$. It is important to observe that the value of $\Delta\mathbf{x}(t) = \beta$ summarizes both the departure and the arrival state of the chain, which will have values -1 and $+1$ in β at the corresponding coordinates, if the states differ. Therefore, the values of $\mathsf{P}\{\Delta\mathbf{x}(u) = \beta|\mathbf{x}(u-)\}$ will just be equal to the respective transitional probabilities of the Markov chain.

Of course, in most financial applications $S(t)$ will denote the logarithm of the asset price. The filtration with respect to which all expectation are taken will be denoted $\mathscr{F}(t)$ and it will represent the information generated by observing this asset price alone, $\mathscr{F}(t) = \sigma\{S(u), 0 \leqslant u \leqslant t\}$.

Process (1) as a switching Lévy process. Suppose that the underlying Markov chain stays at a particular state $\mathbf{x}(t)$ for the time interval (t, u). One can observe that during this period the process $Z(t) = 0$ and the process $S(t) = Y(t) + Z(t)$ is a Lévy process [see Bertoin (1996) for definitions]. Denote the characteristic exponent with $\Psi(\theta, \mathbf{x}(t))$. Since the Markov chain is right continuous, for an N –dimensional ball of radius 1 centered at $\mathbf{x}(t-)$, $\mathsf{Ball}(\mathbf{x}(t-), 1)$, one can always find a [stopping] time $t^\star > t$ such that $\mathbf{x}(u) \in \mathsf{Ball}(\mathbf{x}(t-), 1)$ for all $u \in [t, t^\star)$, or equivalently that $\mathbf{x}(u) = \mathbf{x}(t-)$ for all $u \in [t, t^\star)$, since unit vectors have a distance between them which is higher than unity. Therefore, one can easily conclude that the Markov chain will always be at a constant state in the interval $[t, t^\star)$, and in this case the instantaneous characteristic function will take the form $1 + \varpi\Psi(\theta, \mathbf{x}(t-))\mathrm{d}t + o(\mathrm{d}t)$, with $\varpi = \sqrt{-1}$.

In this context the stochastic process (1) can be thought of a switching Lévy process, inheriting conditionally all the useful properties, namely conditional distributions that are infinitely divisible, or equivalently that their conditional characteristic functions are of the exponential form as given by the Lévy–Khinchine formula.

On the other hand, the [unconditional] univariate process $S(t)$ is not Lévy, in fact it is not even Markov. This is due to the fact that the parameters are dependent

on the chain $\mathbf{x}(t-)$. The fact that the driving process $\mathbf{x}(t-)$ is Markov will prove very useful in the analysis below, and especially in the proof of the main theorem 1.

The observations above indicate that process (1) shares a number of characteristics with a number of models applied into financial data, most prominently it has the same structure as the jump diffusion stochastic volatility specifications that have been very successful in explaining the stylized facts of financial time series and have over performed most other specifications in pricing derivative contracts. One can recognize the volatility variability of (1) courtesy of the hidden Markov chain, the existence of Poisson jumps $J(t)$ with volatility dependent intensity and magnitude and the correlation between volatility changes and changes in $S(t)$ which are due to the effect of process $Z(t)$. Indeed, subsequent sections will formalize these structural similarities and will construct approaches that will exploit them for estimation and pricing purposes.

2 The characteristic function

The main result. The main result of the paper, around which the analysis is built, is summarized in theorem 1 below. Before stating the result, one has to evaluate the conditional characteristic exponent $\Psi_i(\theta)$, as given by the Lévy–Khinchine formula [see Bertoin (1996)]

$$\Psi_i(\theta) = \Psi(\theta, \mathbf{e}_i) = \lambda(\mathbf{e}_i) \int (1 - e^{\varpi\theta\alpha}) \nu_J(\{\mathbf{e}_i\} \times \mathrm{d}\alpha) - \lambda(\mathbf{e}_i) \mathsf{E}\{\nu_J(\mathbf{e}_i)\}. \qquad (2)$$

One also needs to define the characteristic function [not exponent] associated with the measure $\Pi_Z(\{\mathbf{e}_j - \mathbf{e}_i\} \times \cdot)$, namely

$$\Psi_{ji}(\theta) = q_{ji} \int e^{\varpi\theta\alpha} \nu_Z(\{\mathbf{e}_j - \mathbf{e}_i\} \times \mathrm{d}\alpha) - e^{\varpi \sum_{\mathbf{e}_k \in \mathcal{B}} q_{ki} \mathsf{E}\{\nu_Z(\mathbf{e}_k - \mathbf{e}_i)\}}. \qquad (3)$$

Proposition 1 *Consider the process (1) together with assumptions A1–A3. The*

characteristic function of $S(t)$*, given the initial state* $\mathbf{x}(0)$ *will satisfy*

$$\phi(\theta, t|\mathbf{x}(0)) = [\mathbf{x}(0)]^T \exp\{t\mathbf{B}(\theta)\}\iota, \tag{4}$$

where $\mathbf{B}(\theta)$ *has elements of the form*

$$\beta_{ij}(\theta) = \begin{cases} q_{ii} + \Psi_i(\theta), & \text{when } i = j \\ \Psi_{ji}(\theta), & \text{when } i \neq j \end{cases}.$$

Proof: See appendix.

The above system of differential equations (16) can be solved even if some of the conditions imposed on the structure of the stochastic process (1) were not so restrictive. In particular, the system can be solved even if the parameters of the diffusions were explicitly dependent on time, and if the underlying diffusion was not Lévy but mean reverting. In these cases one can follow the same procedure to arrive to a different matrix differential equation, where the matrix $\mathbf{B}$ would be time dependent. The solution of this system would not be of the matrix exponential form, but it can be represented as the product–integral of the matrix $\mathbf{B}$, $\tilde{\mathbf{\Phi}}(\theta, t) = \coprod_{(0,t]}\{\mathbf{I} + \mathbf{B}(\theta, u)\mathrm{d}u\}$ [see Dollard and Friedman (1979) for definitions].

The result of Theorem 1 is very powerful. It allows one to use the closed–form knowledge of the characteristic function in order to estimate the unknown parameters of the model and to price derivative contracts. Estimation can take place by inverting the Fourier transform of the density and computing the likelihood function as in Singleton (1998) [see also Duffie, Pan, and Singleton (1999)], or using method of the moments techniques as in Das (1998). Derivative pricing can be carried out by observing that claims with trigonometric payoffs span the asset space as shown in Bakshi and Madan (2000). The next subsection gives the conditional moments of a process such as (1), which allow a simple implementation of GMM estimation procedures.

The conditional moments. Given the conditional characteristic function of Theorem 1, calculation of the conditional moments can be carried out by simple differentiation. The matrix exponential can be differentiated in the lines of Mathias (1997). Specifically, in order to calculate the first m moments, one has to:

1. Create a sequence of matrices $\{\frac{1}{k!}\mathbf{B}^{(k)}(\theta)\}_{k\in\{0,...,m\}}$, where $\mathbf{B}^{(k)}(\theta)$ has as elements the k–th derivatives of the elements of $\mathbf{B}(\theta)$. Since these elements are given in closed form, so is the sequence $\{\frac{1}{k!}\mathbf{B}^{(k)}(\theta)\}_{k\in\{0,...,m\}}$.

2. Construct the Block Upper Triangular Block Toeplitz [BUTBT] matrix, using as blocks the elements of the matrix sequence above,

$$\tilde{\mathbf{B}}(\theta) = \begin{pmatrix} \frac{1}{0!}\mathbf{B}^{(0)}(\theta) & \frac{1}{1!}\mathbf{B}^{(1)}(\theta) & \cdots & \frac{1}{m!}\mathbf{B}^{(m)}(\theta) \\ \mathbf{0} & \frac{1}{0!}\mathbf{B}^{(0)}(\theta) & \cdots & \frac{1}{(m-1)!}\mathbf{B}^{(m-1)}(\theta) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{0!}\mathbf{B}^{(0)}(\theta) \end{pmatrix}$$

3. The matrix exponential of the above matrix $\exp\{\tilde{\mathbf{B}}(\theta)\}$ will have as block elements the $(N \times N)$ matrices of derivatives $\frac{1}{k!}\frac{\mathrm{d}^k}{\mathrm{d}\theta^k}\exp\{\mathbf{B}(\theta)\}\Big|_{\theta=0}$. In other words, the k–th derivative of the matrix exponential function $\mathbf{B}(\theta)$ will be given by $k!$ times the $(1,k)$ block [with dimensions $(N \times N)$] of the matrix $\exp\{\tilde{\mathbf{B}}(\theta)\}$.

Based on the properties of the characteristic function, the conditional [on the regime $\mathbf{x}(0)$] uncentered moments of the process will be given by simply differentiating (1) at $\theta = 0$, using the matrix exponential differentiating procedure discussed above.

3 Maximum likelihood estimation of continuous time state dependent models

This section discusses the ways in which the results of the previous part can be employed in order to achieve maximum likelihood estimates of a system that obeys (1). This will be based on a discrete finite sample $\{S(0), S(\delta), \ldots, S(n\delta)\}$, where the time interval between successive observations is conventionally taken to be constant, equal to δ.[3]

Construction of the likelihood. Having obtained the characteristic function of (1) in closed form, one can retrieve the conditional [on the state] density of the change $\Delta S(k\delta) = S((k+1)\delta) - S(k\delta)$, for some $k \in \{0, \ldots, n-1\}$. The value of the density at any point ΔS using the Fourier inversion formula [Kendal and Stuart (1977)] will be

$$f(\Delta S|\mathbf{e}_i) = \frac{1}{\pi}\int_0^{\infty} \mathrm{Re}\{e^{-\varpi\theta\Delta S}\phi(\theta, \delta|\mathbf{e}_i)\}\mathrm{d}\theta \tag{5}$$

As discussed in Hamilton (1989) for the discrete-time regime switching model, in the case where the state sequence $\{\mathbf{x}(0), \mathbf{x}(\delta), \ldots, \mathbf{x}((T-1)\delta)\}$ is revealed to the econometrician, the log-likelihood of the sample will be given by

$$\mathscr{L} = \sum_{k=0}^{T-1} \ln f(\Delta S(k\delta)|\mathbf{x}(k\delta)) \tag{6}$$

The econometrician can then maximize the likelihood function (6) numerically, with respect to the parameter vector. In most cases, though, the Markov chain is hidden will have to be filtered out of the sample.

[3] This does not affect the generality of the results, it is straightforward to verify that the results hold if the time interval is variable [but known], equal to δ_i.

The filtering of the unobserved state. An important result in the continuous time Markov chain literature states that if we denote

$$\xi(t|\tau) = \begin{pmatrix} \mathsf{P}\{\mathbf{x}(t) = \mathbf{e}_1|\mathscr{F}(\tau)\} \\ \vdots \\ \mathsf{P}\{\mathbf{x}(t) = \mathbf{e}_N|\mathscr{F}(\tau)\} \end{pmatrix},$$

then conditional distributions of the future states can be formed by computing the matrix exponential of the rate matrix as

$$\xi(t|t+\delta) = \xi(t|t)\exp\{\delta\mathbf{Q}\}$$

Equivalently, one can restate that the transition probability matrix over a period of length δ is equal to the matrix exponential $\exp\{\delta\mathbf{Q}\}$. This result can be applied in order to filter out the unobserved states using the same procedure as in the discrete time filter of Hamilton (1989). In fact Hamilton's filter is applied intact, with two modifications compared to its standard form: (i) The observations are following non-normal distributions, given by the inversion (5), and (ii) the transition probabilities are given by the matrix exponential $\exp\{\delta\mathbf{Q}\}$. Given the parameter values, the likelihood can be computed using the following procedure[4]

1. Compute the transition probability function $\Pi = \exp\{\delta\mathbf{Q}\}$;
2. Create the $(T \times N)$ matrix $\mathbf{H}$ of density evaluations of the sample over the states, with (k, i) element equal to $f(\Delta S(k\delta)|\mathbf{e}_i)$;
3. Loop over the sample [counting $k \in \{0, \ldots, T-1\}$] and compute the following
   (a) Multiply the elements of the k-th row of $\mathbf{H}$ with the elements of the

[4] $\odot$ denotes element-by-element matrix multiplication and $\div$ denotes element-by-element matrix division.

vector[5] $\xi(k\delta|(k-1)\delta)$, creating the vector $\zeta = \xi(k\delta|(k-1)\delta) \odot \mathbf{H}_k$;

(b) Compute the probabilities $\xi(k\delta|k\delta) = \zeta \div \{\mathbf{1}^\top \zeta\}$;

(c) Create the new forecast vector $\xi((k+1)\delta|k\delta) = \xi(k\delta|k\delta)\mathbf{\Pi}$;

4. The log-likelihood will be given by

$$\mathscr{L} = \sum_{k=0}^{T-1} [\ln \xi(k\delta|k\delta)]^\top \mathbf{H}_k \tag{7}$$

Details of the procedure and proofs can be found in Hamilton (1994). Again, the econometrician can numerically maximize the above log-likelihood with respect to the vector of parameters. As a natural byproduct of the log-likelihood computation procedure described above, one retrieves the very important series of vectors $\xi(k\delta|k\delta)$, which reveal the distribution of the unobserved state, conditional on the information which prevailed at the time.

Speeding up the computations. The number of density evaluations in order to compute one value of the likelihood function is equal to the number of elements of the matrix $\mathbf{H}$, namely TN. Following (5) each density computation requires one numerical inversion of the characteristic function, or equivalently one numerical integration. When the sample size and/or the number of possible states is large this might pose computational difficulties, considering the computing resources demanded by numerical integration routines.

If the case where the problems stem from a large sample size, one natural density approximation method would be to reduce the number of density evaluations by using a finite grid and interpolate. Instead of TN density evaluations only $\tilde{T}N$ are carried out, with $\tilde{T}$ being the grid size. In order to ensure the non-negativity of

[5] In order to start the algorithm one needs to set the value of the vector $\xi(0|-\delta)$. This can be either set to the ergodic distribution of the Markov chain, given by the limit $\lim_{t\to\infty} \exp\{t\mathbf{Q}\}$, or it can be estimated amongst the other parameters. See Hamilton (1994) for details.

the density, the log-spline approximation methodology of Kooperberg and Stone (1991) can be used, where a B-cubic spline is used in order to interpolate between the various grid points.

On the other hand, the closed form of the conditional moments can be utilized in order to produce density approximations and expansions. Expansions around the normal [or other densities] include the Hermite, Edgeworth and Gram Charlier expansions [see Kendal and Stuart (1977) for details] and the Hermite based SNP expansions of Galland and Nychka (1987).

Expansions around the normal can prove very useful and accurate if the conditional density are of moderate skewness and kurtosis. As pointed out in Jondeau and Rockinger (1998), higher moments that indicate significant departures from normality can cause the density function to take negative values, rendering it inapplicable in the context of likelihood estimation. In contrast, the SNP methods ensure positivity, but are computationally burdensome when it comes to moment computations.

An interesting candidate is the Pearson IV family of densities, designed to approximate strongly leptokurtic unimodal densities, which are frequently encounter in the financial data sets. The density can be expressed in terms of the skewness and kurtosis coefficients [see Kendal and Stuart (1977) for implementation details].

4 Risk neutral derivative pricing

In the previous section, the characteristic function of a process that depends explicitly on a continuous time Markov chain was utilized for the purpose of forming maximum likelihood estimators. This section turns to the issue of pricing derivative contracts of the European type.

There are two general contingent claim pricing issues that are dealt with below. The first regards the formal construction of the risk adjusted probability measure,

under which all discounted assets form martingales. It is shown how this measure is generated as a modification of the physical measure, where the various parameter adjustments reflect the effects of different scenarios on the marginal utility of the representative agent. One further assumption is needed for the derivation of the results:

A4 The utility of the representative agent depends on the state of the economy, and *not* on the distribution of this state.[6]

The risk neutral measure. The process of the asset log-price under the risk neutral measure, in line with Merton (1976) and Bates (1995), will obey the stochastic differential equation

$$S(t) = s + \bar{Y}(t) + \bar{J}(t) + \bar{Z}(t), \tag{8}$$

where the bars denote quantities under the risk adjusted probability measure. The diffusion part $\bar{Y}(t)$ will satisfy by construction [with r the constant interest rate]

$$\bar{Y}(t) = rt + \int_0^t \sigma(\mathbf{x}(u-))\, \mathrm{d}\bar{W}(u)$$

The parameters that dictate these risk neutral adjustment for the other processes $\bar{J}(t)$ and $\bar{Z}(t)$ can be constructed in the fashion described below.

Representative agent general equilibrium models derive the risk neutral parameters as modifications of the true ones, taking into account the effect of the state variables on the diffusion of the percentage changes of the marginal utility of nominal wealth $U_W(t)$. Let the diffusion of these percentage changes be denoted by $\mathrm{d}\mathcal{M}(t) = \frac{\mathrm{d}U_W(t)}{U_W(t)}$. Denote by $\Delta\mathcal{M}_{ij}(t)$ the former quantity, given a chain change

[6]Alternatively, the agents do not exhibit "belief dependent" utility, in the sense of Veronesi (2001). In the setting of Veronesi (2001) the utility depends on a measure of the uncertainty which surrounds the state. The higher the dispersion of this uncertainty, the lower the utility.

from state $\mathbf{e}_i$ to $\mathbf{e}_j$, and with $\Delta\mathcal{M}_i(t)$ the same quantity given that $\mathbf{x}(t) = \mathbf{e}_i$. The relationships between the true and the risk neutral parameters follow from the relationship between the true and the risk neutral probability measure, and the rôle of $\Delta\mathcal{M}(t)$ as the state price density or the Radon-Nikodym derivative of the risk neutral probability measure with respect to the true probability measure. For convenience, and without loss of generality fix the initial time at zero, with $\mathcal{M}(0) = 0$ and consider a process $\Lambda(t)$ with $\Lambda(0) = 0$. The expectation under the risk neutral probability measure of $\Lambda(t)$ conditional on $\mathscr{F}(0)$, can be written as

$$\bar{\mathsf{E}}\{\Lambda(t)|\mathscr{F}(0)\} = \mathsf{E}\{(1 + \mathcal{M}(t))\Lambda(t)|\mathscr{F}(0)\}.$$

A series of lemmas below discusses the implications on the risk adjusted parameters.[7]

Lemma 2 *Under assumption (A4) the filtered physical probabilities are equal to the risk adjusted ones,*

$$\bar{\xi}(t|t) = \xi(t|t)$$

Proof: See appendix.

In all empirical work on option pricing with stochastic volatility, the volatility is considered observed and equal to its filtered value. Apparently this might cause misspecification errors, if the variability of these estimators is high enough. There are two sources of such errors: (i) The volatility process is not directly observed and in fact has a [discrete] distribution, and (ii) this distribution will take a different form under risk neutrality. The first misspecification source is discussed in

[7] Similar results to lemmas 3 and 4 are given in Bates (1995) for the jump diffusion process. Jarrow, Lando, and Turnbull (1997) use a similar construction of the risk adjusted measure when they consider credit derivatives, with the states of the Markov chain representing the various possible credit ratings.

the empirical part of this paper, where the actual variability of the filtered probabilities is presented. Assumption (A4) and the resulting result of lemma (2) give the condition that has to be satisfied in order for one to ignore the second source, namely the risk adjustment. In an asset pricing exercise Veronesi (2001) identifies the risk adjusted filtered probabilities as

$$\bar{\xi}_j(t) = \frac{\xi_j(t) U_W(t|\mathbf{e}_j)}{\sum_i \xi_i(t) U_W(t|\mathbf{e}_i)}$$

when the utility function exhibits constant relative risk aversion. Further discussion on this problem is left for future research.[8]

Lemma 3 *The rate matrix under risk neutrality will have elements of the form*

$$\begin{aligned} \bar{q}_{ji} &= q_{ji} \mathsf{E}_t \left\{ 1 + \Delta \mathcal{M}_{ji}(t) \right\}, \text{ for } i \neq j, \text{ and} \\ \bar{q}_{ii} &= -\sum_{j=1}^{N} \bar{q}_{ji} \end{aligned} \tag{9}$$

Proof: See appendix

Lemma 4 *Denote with* $\mu_{ji} = \mathsf{E}\{\nu_Z\{\mathbf{e}_j - \mathbf{e}_i\}\}$ *the expected jump size when a regime switch from state* $\mathbf{e}_i$ *to* $\mathbf{e}_j$ *takes place [*$\bar{\mu}_{ji}$ *the corresponding quantity under risk neutrality]. Then*

$$\bar{\mu}_{ji} = \mu_{ji} + \frac{\mathsf{Cov}\left\{\nu_Z\{\mathbf{e}_j - \mathbf{e}_i\}, \Delta \mathcal{M}_{ji}\right\}}{\mathsf{E}\left\{1 + \Delta \mathcal{M}_{ji}\right\}}. \tag{10}$$

Proof: See appendix.

[8] For example, one might want to impose a utility form which is decreasing with respect to the variance of the filtered volatility process $\mathsf{Var}\{\xi(t|t)\}$. Intuitively in such a setting agents the higher the certainty of the state, the happier agents are. This should imply that under risk neutrality higher volatility states will have higher risk neutral probabilities.

Lemma 5 *Deterministic jumps as the ones in Naik (1993) or the correlation adjustments of the next section do not carry any price of risk.*

PROOF: Such jumps do not have any covariance with the process $\Delta\mathcal{M}_{ji}(t)$.

There are two important issues that need to be addressed concerning equation (8) and the results of lemmas 2, 3 and 4:

First, the relationships between the *true* process given in equation (1) and the *risk neutral* process given in (8). The instantaneous drift of the true process has disappeared; moreover the risk neutral process does not exhibit mean reversion through the mean reverting behavior of the CTMC $\mathbf{x}$. One could argue that it might not be necessary to include such a parametrized drift in the first place. This is not true: Lo and Wang (1995) show that although the drift does not enter the option pricing formulas directly, it can have a substantial effect on the prices. The intuition is straightforward: if the true model is estimated correctly then the estimated parameters that enter the formulae directly [the volatilities for example] will be different compared to the ones estimated using the wrong specification.

Second, the computation of the prices of risk in equations (9) and (10). Given the information of the representative agent, $\mathscr{F}(t)$, the securities market is incomplete, that is to say the risk neutral parameters cannot be retrieved and therefore correct option prices cannot be computed. There are three possible ways of tackling this problem:

1. One can make the assumption that the regime risk and the jump risk cannot be diversified, or equivalently that the expected marginal utility of wealth is not affected by regime changes, giving $\mathsf{d}\mathcal{M}_{ij} = 0$. This somewhat strict assumption results into risk neutral parameters that are equal to their true counterparts, a methodology utilized in the early approaches of derivative pricing.

2. Alternatively, one can assume a parametric form of the prices of risk, for example that $\mathsf{E}\left\{\mathsf{d}\mathcal{M}_{ij}\right\} = \varphi_{ij}$, a constant. In addition, one has to augment the

representative agent's information with the prices of the options written in the past, that is to say $\tilde{\mathscr{F}}(t) = \mathscr{F}(t) \otimes \sigma \{C(s,.), s \leqslant t\}$, in order for those extra parameters to be estimated. Recent studies that adopt this methodology include Benzoni (1999), Chernov and Ghysels (2000) and Pan (2002)

3. Finally, one can use historical estimates of the regime impact on the marginal utility of wealth in a general equilibrium framework, and use equations (9) and (10) directly. This approach has not been utilized directly in the option pricing literature, mainly due to the failure of general equilibrium models to explain excess returns and equity volatilities [see Cohrane (1997)]. The work of Jackwerth and Rubinstein (1996) and Jackwerth (1999) sheds some light to the sources of incompatibility between realized returns and the corresponding option prices.

The European call option formula. As shown in the previous paragraph the functional form of the stochastic process under the risk adjusted probability measure remains fundamentally the same, although the parameter values are adjusted in order to accommodate the fears of the financial markets. Nevertheless, a form of equation (1) of theorem 1 on page 9 still holds for the asset return distribution after the risk adjustments have taken place. The characteristic function of this distribution is denoted with $\bar{\phi}(\theta, t|\mathbf{x}(0))$. The price of a European call option, maturing after time τ, will be given by the expectation of its payoffs, under the risk neutral measure, that is to say

$$C(\tau, K; \theta) = e^{-rt} \bar{\mathsf{E}} \{[S(\tau) - K]^+ | \mathscr{F}(0)\} = e^{-rt} \bar{\mathsf{E}}_0 \{S(\tau) - K\}^+, \qquad (11)$$

where $\bar{\mathsf{E}}\{\diamond\}$ denotes the expectation taken under the risk neutral measure, and the vector θ includes all the parameters estimated at time $t = 0$, as well as the initial state $\mathbf{x}(0)$ and price $S(0)$. The above implies the following

Proposition 6 (European call option price) *Consider an asset whose log-price obeys the stochastic differential equation (1). Assume that the price of the state shift risk and the price of the jump risk affect the estimated parameters as in equations (9) and (10). Then, the price of the European call, maturing after time τ with strike price equal to K, will be given by*

$$C(\tau, K; \theta) = S(0)\,\Pi_1(\tau, K; \theta) - e^{-r\tau} K \Pi_2(\tau, K; \theta), \tag{12}$$

where

$$\Pi_n(\tau, K; \theta) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \mathsf{Re}\left\{ \frac{e^{-\varpi\theta \ln[K/S(0)]} \bar{\phi}_n(\tau, \theta; \vartheta)}{\varpi\theta} \right\} \mathrm{d}\theta,\ n = 1, 2$$

and the following quantities are defined:

$$\bar{\phi}_2(\tau, \theta; \vartheta) = \bar{\phi}(\theta, \tau|\mathbf{x}(0))$$

$$\bar{\phi}_1(\tau, \theta; \vartheta) = \frac{\bar{\phi}(\theta - \varpi, \tau|\mathbf{x}(0))}{\bar{\phi}(-\varpi, \tau|\mathbf{x}(0))}$$

and the Fourier transform is given as in equation (1)

$$\bar{\phi}(\theta, \tau|\mathbf{x}(0)) = e^{\varpi\theta r\tau} [\xi(0)]^{\mathsf{T}} \exp\left\{\tau \bar{\mathbf{B}}(\theta)\right\} \iota.$$

Proof: Follows Bakshi and Madan (2000, Case 2).

5 Connections with stochastic volatility models

The previous section of this paper has exclusively and extensively analyzed various aspects of processes that explicitly depend on continuous time Markov chains. This section attempts to bridge the gap between such processes and ones that exhibit diffusive volatility, with and without jumps. To that end, it will try and explore how one can use a sequence of Markov chains to approximate a diffusion.

The intuition behind this approach is that if one constructs state spaces that are sequentially denser, and if the appropriate transition probability structure is maintained, then the sample paths of resulting Markov chain will approach the sample paths of diffusions with prescribed drift and volatility parameters.

Such an approach can be beneficial for numerous applications: Firstly, using such approximations, one can carry out maximum likelihood estimation of the stochastic volatility parameters following the results of section 3 on page 12, without having to resort to simulation based methods. Secondly, one can use the results of section 4 and in particular equation (12) in order to price European options, in cases where semi-closed form solutions are not available. In such cases, for example if the volatility follows a log-diffusion, one should normally engage into numerically solving systems of PDEs.

The plan of this section is the following: The approximation schemes for the simplest univariate diffusion which is uncorrelated with the price process are initially introduced. Subsequently, the necessary modifications are discussed, which allow one to include correlations and jumps that have intensities and distributions that depend on the latent process. A brief overview of the methods that would allow jumping volatility and the multivariate latent variable case follows.

The zero–correlation case. The objective of this paragraph is to find the appropriate family of Markov chains that would approximate the volatility diffusion of the stochastic differential equation

$$\begin{aligned} \mathrm{d}S(t) &= \mu(\sigma(t-))\mathrm{d}t + \sigma(\eta(t-))\mathrm{d}W(t) \\ \mathrm{d}\eta(t) &= \alpha(\eta(t-))\mathrm{d}t + \beta(\eta(t-))\mathrm{d}V(t) \end{aligned}$$

where the innovations $W(t)$ and $V(t)$ are assumed to be [for the time being] independent Wiener processes. Such a specification includes models that have been

widely used in the literature such as the Heston model:

$$\sigma : \eta \to \eta$$
$$\alpha : \eta \to \theta(\bar{\sigma} - \eta)$$
$$\beta : \eta \to \varphi\sqrt{\eta}$$

and the log-volatility model:

$$\sigma : \eta \to \exp\{\eta\}$$
$$\alpha : \eta \to \theta(\bar{\sigma} - \eta)$$
$$\beta : \eta \to \varphi$$

Markov chain approximations to diffusions are not a new technique: A series of research papers and books summarized in Kushner (1990) and Dupuis and Kushner (2001) have developed a number of optimal methods, applied mainly in a stochastic control framework.[9] The main idea behind the convergence, and the very mild condition that has to be satisfied, is the one coined "local consistency", where the instantaneous drift and volatility of the continuous time Markov chain match [at least asymptotically] the ones of the diffusion process in question. As noted in Dupuis (2002), "One of the key advantages of the Markov chain approach is that this is done using purely probabilistic methods, and consequently far less regularity is required of the problem data". In general, the local consistency argument can be formalized in the following

Definition 7 (Local Consistency) *For each $h > 0$ define a grid of equidistant*[10]

[9]Examples of financial applications include Frey and Runggaldier (2001) on the filtering of high frequency discrete volatility, Laurent and Leisen (2000) on option pricing when the price itself is approximated using Markov chains and Chiarella, Pasquali, and Runggaldier (2001) on the filtering of the term structure of interest rates.

[10]*The points need not be equidistant. Here the equal distance is used for ease of the exposition, and since without loss of generality nonlinearities can be introduced by the function $\sigma(\cdot)$.*

points $\mathcal{G}^h$ which tends to cover the domain of η as $h \downdownarrows 0$. Denote the elements of the grid $\mathcal{G}^h = \{\eta_1, \eta_2, \ldots, \eta_n\}$ and let $\mathbf{x}^h(t)$ be a Markov chain that lives in $\mathbb{R}^n$ with corresponding rate matrix $\mathbf{Q}^h$. The associated filtration is generated as $\mathcal{F}^h(t) = \sigma\{\mathbf{x}^h(u), 0 \leqslant u \leqslant t\}$. Let E_t^h denoted the expectation taken conditional on this filtration. Define naturally the Markov chain $\eta^h(t) = [\eta^h]^\intercal \mathbf{x}^h(t)$ that lives in the grid $\mathcal{G}^h$. The Markov chain will approximate the diffusion with infinitesimal drift $\alpha(\cdot)$ and volatility $\beta(\cdot)$ if the functions $\alpha(\cdot)$ and $\beta(\cdot)$ are Lipschitz continuous, and for some $\varrho > 0$ and $\delta > 0$, with $\Delta\eta^h(t) = \eta^h(t+\delta) - \eta^h(t)$

$$\mathsf{E}_t^h\{\Delta\eta^h(t)\} = \alpha(\eta^h(t))\delta + O(h^\varrho\delta) \tag{LC1}$$

$$\mathsf{E}_t^h\{\Delta\eta^h(t) - \mathsf{E}_t^h\{\Delta\eta^h(t)\}\}^2 = \beta(\eta^h(t))\delta + O(h^\varrho\delta) \tag{LC2}$$

$$|\Delta\eta^h(t)| = O(h) \tag{LC3}$$

The approximation scheme employed here is the same as the one in Piccioni (1987) [see also Kushner (1977) for further details]. The rate matrix $\mathbf{Q}^h$ is of the Jacobi form[11], with elements given by

$$q^h(\eta_i, \eta_{i-1}) = \frac{1}{2h^2}\beta(\eta_i) + \frac{1}{h}\alpha^-(\eta_i) \tag{13}$$

$$q^h(\eta_i, \eta_i) = -\frac{1}{h^2}\beta(\eta_i) - \frac{1}{h}|\alpha(\eta_i)| \tag{13′}$$

$$q^h(\eta_i, \eta_{i+1}) = \frac{1}{2h^2}\beta(\eta_i) + \frac{1}{h}\alpha^+(\eta_i) \tag{13″}$$

for all interior states, while the first and last [n-th] state are made absorbent. In order to be consistent with the notation of section 1, one can construct the vector η^h with elements $\{\eta_i\}_{i=1}^n$, and define the chain $\mathbf{x}^h(t)$ which obeys the rate matrix $\mathbf{Q}^h$ [with the elements described in (13)]. Then of course the process $\eta(t)$ is approximated by $\eta^h(t) = [\eta^h]^\intercal \mathbf{x}^h(t)$.

Since the constructed rate matrix is of the Jacobi form, permitted transitions

[11] $\beta^\pm(\cdot)$ denotes the positive/negative part of a function.

are only towards adjacent states. This is a unique feature of diffusion processes, which allows one to approximate them in a very parsimonious way.[12] One can further note that the drift and volatility of the Markov chain process are

$$\alpha^h(\eta_i) = \alpha(\eta_i)$$
$$\beta^h(\eta_i) = \beta(\eta_i) + h|\alpha(\eta_i)| = \beta(\eta_i) + O(h)$$

satisfying conditions (LC1)-(LC2). Condition (LC3) can be trivially satisfied by the appropriate choice of the grid $\mathcal{G}^h$.

Correlation adjustments. The introduction of correlation will be achieved by adding deterministic jumps to the price process whenever the volatility process changes state. The size of these changes will dictate the correlation $\mathsf{Cor}(dW(t), dV(t)) = \rho(\eta(t))dt$. The process can be written as

$$\begin{aligned} \mathrm{d}S(t) &= \mu(\eta(t-))\mathrm{d}t + \sqrt{1-\rho^2(\eta(t-))}\sigma(\eta(t-))\mathrm{d}W(t) \\ &\qquad + \rho(\eta(t))\sigma(\eta(t-))\mathrm{d}V(t) \\ \mathrm{d}\eta(t) &= \alpha(\eta(t-))\mathrm{d}t + \beta(\eta(t-))\mathrm{d}V(t) \end{aligned}$$

where the first equation by substituting $\mathrm{d}V(t)$ yields

$$\begin{aligned} \mathrm{d}S(t) = \mu(\eta(t-))\mathrm{d}t &+ \sqrt{1-\rho^2(\eta(t-))}\sigma(\eta(t-))\mathrm{d}W(t) \\ &+ \rho(\eta(t))\frac{\sigma(\eta(t-))}{\beta(\eta(t-))}[\mathrm{d}\eta(t) - \alpha(\eta(t-))\mathrm{d}t] \end{aligned}$$

For the approximating chain, this will dictate the addition of the component $\mathrm{d}Z^h(t) = \mathrm{d}Z(\eta^h(t))$ in the following fashion, considering that the continuous time Markov

[12] Indeed, this feature has been exploited in order to test the assumption of whether or not a process is a diffusion. See Aït-Sahalia (2002) for more details. Such a property will not be satisfied in the sequel, where models that incorporate jump diffusive volatility are discussed.

chain takes the value $\eta^h(t) = \eta^h$ at time t:

$$\begin{aligned}
\mathrm{d}S(t) &= \mu(\eta^h)\mathrm{d}t + \sqrt{1-\rho^2(\eta^h)}\sigma(\eta^h)\mathrm{d}W(t) + \mathrm{d}Z^h(t) \\
\mathrm{d}Z(t) &= \pm h\rho(\eta^h)\frac{\sigma(\eta^h)}{\beta(\eta^h)}, \\
&\qquad \text{when the chain changes state towards } \eta^h \pm h
\end{aligned}$$

These probability changes take place with the probabilities given in equation (13). The compensator of the process is

$$\rho(\eta)\alpha(\eta)\frac{\sigma(\eta)}{\beta(\eta)}\mathrm{d}t$$

Processes with path-dependent jumps. The inclusion of jumps is straightforward, compared to the correlation adjustments discussed in the previous paragraphs. One element worth noting is the ability of the approximating chain model to incorporate jumps that have not only intensities, but also densities that depend on the level of the latent process. The jump process $\mathrm{d}J^h(t) = \mathrm{d}J(\eta^h(t))$ can be constructed as

$$\mathrm{d}J^h(t) = \int_A \Pi_J^h(\eta^h \times \mathrm{d}\alpha) - \lambda^h(\eta^h)\mathsf{E}\{\nu_J^h\}\mathrm{d}t$$

where the sequences $\Pi^h \to \Pi$ [or $\nu^h \to \nu$] and $\lambda^h \to \lambda$ as $h \downdownarrows 0$. Such sequences can be easily constructed as

$$\begin{aligned}
\Pi_J^h(\eta^h \times \mathrm{d}\alpha) &= \Pi_J(\mathbf{x}^h(t) \times \mathrm{d}\alpha) \\
\nu_J^h(\eta^h) &= \nu_J(\mathbf{x}^h(t)) \\
\lambda_J^h(\eta^h) &= \lambda_J(\mathbf{x}^h(t))
\end{aligned}$$

Jumping volatility, feedback effects and multivariate models. There has been a growing strand in the literature that exploits processes where the volatility itself is allowed to exhibit jumps of random magnitudes. Such processes have been ei-

ther estimated using simulated methods as in Eraker, Johannes, and Polson (2000), or they are setup in a discrete time setting, and therefore are estimated through ARCH-type methods, as in Pan (1997). The generic approach presented here gives the flexibility to the researcher to approximate a latent diffusion with an appropriate Markov chain, even in the case where the diffusion exhibits discontinuities. Such methods are introduced and discussed in Kushner and DiMasi (1978). Intuitively, a jump-diffusion process will be approximated by a Markov chain, where the transition matrix will allow movements towards states that do not neighbor. The transition probabilities towards these non-adjacent states will be reflecting the intensity to jump as well as the jump distribution.

Another class of models which could be encompassed in this generic specification, albeit after a number of modifications, consists of models where the asset price [or the return] itself enters the latent variable equation. To approach such models, it suffices to observe that the the rate matrix elements in equations (13) will depend on the observed process. The Markov chain in such a case will not be a homogeneous one, and the characteristic function, will not be given by the matrix exponential as in theorem 1. Instead, one has to utilize the product-integral as noted in page 10.

Multivariate factor models have been used recently to model equity behavior, with the papers of Chernov, Gallant, Ghysels, and Tauchen (1999) and Chernov, Gallant, Ghysels, and Tauchen (2002) giving some examples, together with an overview of the EMM estimation procedure which is employed. In the framework discussed here, such approaches can be easily accomodated, following the univariate case paradigm and the results discussed, for example, in Kushner (1990). When the multivariate diffusion lives in $\mathbb{R}^N$, the construction of the approximate state-space grid can be constructed by "triangulation"; for example in the bivariate case each point of the state-space is connected with three other points.

6 Estimation of continuous state space models

This section applies the results of the previous parts, and attempts to estimate a stochastic log-volatility model with jumps, similar to the one discussed in Benzoni (1999). The choice of this specification is based on two factors: First such a model has been used extensively to model volatility series, in the standard stochastic volatility form as well as in its EGARCH approximation form. Second and foremost, option prices of such a model are *not* available in closed form, something that allows one to explore the methods analyzed here to their full potential.

Daily data on the SP500 index, spanning ten years from April 1987 to December 1997 were used. The choice of the index is based on its popularity amongst academics and practitioners alike as a market proxy, and on the fact that it underlies the SPX options, which are the most liquid option contracts worldwide. The size of the sample is chosen in order to include the crash of October 1987 in order to examine the behavior of the estimating process. The choice of period after 1987 is motivated by the results of Bates (1997), and in particular by the documented change of the behavior in the options market after the crash. The raw data were subsequently transformed into the log-return series used in the estimation procedure. No prefiltering or seasonality adjustments were carried out as in Andersen, Benzoni, and Lund (1998), since this would destroy features that might prove significant in option pricing [see also Chernov, Gallant, Ghysels, and Tauchen (1999) for similar arguments]. Although option pricing using the approach presented here is not carried out, the choices above have been made keeping in mind the compatibility of the results with future research that uses both index and option price series.

The models considered here are all nested in the specification

$$\begin{aligned}
\mathrm{d}S(t) &= \left(\mu - \frac{\sigma^2(t-)}{2} - \lambda\sigma(t-)\mu_J\right)\mathrm{d}t + \sigma^2(t-)\mathrm{d}W(t) + \lambda\sigma(t-)\mathrm{d}J(t) \qquad (14)\\
\mathrm{d}\ln\sigma^2(t) &= \theta\left(\ln\bar{\sigma}^2 - \ln\sigma^2(t-)\right)\mathrm{d}t + \phi\sqrt{\sigma^2(t-)}\mathrm{d}V(t)\\
\mathrm{d}J(t) &\sim \mathcal{N}\left(\mu_J\sigma(t-) - \frac{\sigma_J^2}{2}, \sigma_J^2\right)
\end{aligned}$$

with $\mathrm{Corr}(\mathrm{d}W(t), \mathrm{d}V(t)) = \rho\mathrm{d}t$

[Table 1 about here.]

Three models are estimated, and the results are presented in table 1. The first model, denoted "SV" is the simple stochastic volatility model without jumps, where the volatility shocks are independent to the shocks on the asset. "SVρ" denotes the model augmented with a constant correlation between the two Wiener processes $\mathrm{d}W(t)$ and $\mathrm{d}V(t)$. "SVρJ" is the full jump diffusion stochastic volatility specification. Following recent evidence [see for example Pan (2002) for a similar analysis] the jump intensity is not constant, but is proportional to the volatility of the underlying process. A novel feature of model (14) is that the jump itself has a distribution that explicitly depends on the latent process. In particular, the mean jump [in percentage terms] is also proportional to $\sigma(t-)$.

The grid construction. The grid is constructed by dividing the interval $[-7, 1]$ in 40 subintervals of equal length 0.2. This is the interval where $\ln\sigma^2(t)$ lies, implying that $\sigma(t) \in [3\%, 163\%]$ p.a.[13] The end intervals have been chosen in such a way as to ensure that the filtered probabilities of these states over the sample is negligible. In that fashion the fact that the support of the volatility process is a closed interval rather than the half line has insignificant consequences on the pa-

[13]The very high end value is courtesy of the "abnormal" market behavior during the '87 crash period. Nevertheless, the estimation algorithm was very successful in dealing with this subperiod.

rameter estimates. Figure 1 shows the volatility path across time, illustrating this point . The grid intensity is chosen in such a way as to ensure that the likelihood value does not exhibit any further sensitivity with respect to further grid subdivisions. This procedure, although intuitive, is somewhat informal. More formal grid selection procedures are an interesting topic for further research.

Estimation results. The results of the maximum likelihood estimation for the three models are reported in table 1. The unit time interval is assumed to be one year, rendering all results in *per annum* terms. The estimated parameters are in line with other studies, carried out mainly using the EMM procedure. The reported standard errors are computed by numerically inverting the Hessian matrix, which is approximated by perturbating the likelihood function by 0.1% of its arguments. Based on these standard errors one can conclude that all parameters are strongly significant, for all standard significance levels.

The inclusion of the correlation verifies [in a qualitative sense] the results of Andersen, Benzoni, and Lund (1998, Tables IV and VI) who also use a log-variance specification for the volatility, but carry out the estimation of the parameters using the simulation based EMM. They are also related to the results of Pan (2002, Table 1). In particular, after the correlation [which has a magnitude of around -0.6] has been accounted for, the strength of the volatility mean reversion decreases. The inclusion of jumps is estimated to have a similar impact in the aforementioned papers and in this one,[14] namely the intensity of jumps occurrences [average 1.70 jumps per year, up to about 8 jumps per year during the '87 crash episode], the negative expected jump [average jump size -0.58%, down to -3% in '87] and the jump variance [with standard deviation 2.25%].

[Figure 1 about here.]

[14]The figures are based on the average estimated volatility without the jump component which is around 12% p.a.

As noted before, one very important byproduct of the estimation procedure employed here is the series $\{\xi(t|t)\}$, resembling the distribution of the latent volatility process, conditional on the information that prevailed at the time, $\mathscr{F}(t)$. Based on that discrete distribution, one can easily compute its expectation, which will play the rôle of the filtered volatility which is displayed in figure 1. It is very significant to note that no further reprojection is needed, as is the case when textscemm estimators are constructed. In addition, since one obtains the whole *distribution* of the unobserved process, confidence intervals of the filtered estimators can be easily constructed. This discussion is postponed until the next paragraph which deals with exactly this feature.

[Figure 2 about here.]

The volatility transition kernels. A second byproduct of the estimation procedure discussed here, is the approximation of the volatility transition kernels, based on the estimated transition probabilities for the Markov chain. These transition kernels are displayed in figure 2. They are computed by simply taking the matrix exponential $\exp\{\tau\mathbf{Q}^h\}$, with τ the horizon in question. One can easily observe the change in the shape of the transition kernels, as the time horizon increases from one week to one year. When the volatility after a year is concerned, the different densities are virtually indistinguishable, for the given estimates of the mean reversion parameter.

[Figure 3 about here.]

[Figure 4 about here.]

The variability of the volatility The distribution of the discretized volatility can be very useful for the researcher in order to assess the uncertainty associated with the estimator through time. Such an investigation is the first [to the author's

knowledge] in the stochastic volatility literature.[15] Figure 3 is used to carry out this investigation, illustrating the empirical findings for the whole sample. Subfigure (a) displays the filtered value of the volatility process across time $\hat{\sigma}(t)$, computed as the weighted average

$$\mathsf{E}_\sigma(t) = \hat{\sigma}(t) = \sum_i \xi_i(t|t)\sigma(\mathbf{e}_i).$$

Note that the volatility in figure 3 does not include the contribution of the jump component, and therefore is somewhat different than figure 1.

In order to asses the relative uncertainty regarding the volatility level, it is convenient to introduce the $[5\%, 95\%]$ spread interval of this filtered estimator, computed by constructing the cumulative density functions based on $\xi(t|t)$ and interpolating. This spread is denoted $\mathsf{SP}_\sigma(t)$ for future reference. The relative spread is constructed as the ratio $\mathsf{RSP}_\sigma(t) = \frac{\mathsf{E}_\sigma(t)}{\mathsf{SP}_\sigma(t)}$. This construction is better illustrated in subfigure (b). It is obvious that low volatility periods are accompanied by higher relative uncertainty. A further, more formal discussion of this point, is postponed to a later paragraph that deals with exactly that information.

Figure 4 magnifies the period 1994-1996 in order to clearly illustrate the changes of beliefs. Subfigure (a) gives the cumulative densities of the filtered volatility. These densities have been constructed from the vectors $\xi(t|t)$ and then applying linear interpolation. One can clearly observe the shifts of the agents' beliefs concerning the volatility level. Subfigure (b) displays the filtered volatility accompanied with the spread for the shorter period.

[Table 2 about here.]

As noted before, one straightforward observation is that the relative spread of the estimator $\hat{\sigma}(t)$ in figure 3(b) is not constant across time, but exhibits cyclicalities. A second observation is that these cyclicalities follow the volatility cycles

[15] Veronesi (2001) discusses a similar case when calibrating an asset pricing model of agents with belief dependent utility.

presented in figure 3(a). Table 2 attempts to offer a more detailed and formal verdict. Four nested regressions are run, based on the general form

$$\Delta \mathsf{E}_{\xi}(t) = a_0 + a_1 \mathsf{E}_{\xi}(t) + a_2 \mathsf{RSP}_{\xi}(t) + a_3 \Delta \mathsf{RSP}_{\xi}(t)$$

$$\text{Model A} : a_2 = a_3 = 0$$

$$\text{Model B} : a_2 = a_4 = 0$$

$$\text{Model C} : a_3 = a_4 = 0$$

$$\text{Model D} : a_3 = 0$$

The unit root hypothesis is tested and rejected [as expected] in regression (A). Regression (B) tries to draw inference on the relationship between expectation changes and the relative uncertainty *level*, as proxied by $\mathsf{RSP}_{\sigma}(t)$. Both models seem to have identical performance based on the reported mean squared error. Models (C) and (D) link expectation changes and uncertainty *changes*. It is apparent that there is a very strong relationship between these two variables, with a superior mean squared error compared to the previous regressions. It can be therefore safely concluded that economic agents seem to have a higher degree of certainty concerning high volatility episodes.

These findings could have significant implications for option pricing purposes. Since quiet market periods are coupled with high relative uncertainty over the volatility level, considering the filtered volatility as an observed variable can easily lead to pricing errors. The nonlinear structure of option prices demands that the unobserved volatility has to be integrated out for the correct option price to be computed. Proposition 6 offers a convenient way of approximating this procedure arbitrarily closely.

On the other hand, risk adjustments of volatility averse agents should create premia that affect the distribution of the unobserved volatility in an asymmetric fashion. In general, due to the risk aversion, adverse event are more probable in the risk neutral world compared to their true probabilities. Thus, during the

periods where the estimated volatility is low but its variance is high these premia could well have a substantial impact on the prices of derivative contracts. This very important issue can be resolved by parameterizing the utility function in the appropriate way. Such an approach is left for further research, where option prices along with the return data are employed for the identification of the model.

7 Conclusions

This paper is an attempt to bridge the gap between the discrete state latent variable models introduced in Hamilton (1989) and their continuous state counterparts discussed in Ghysels, Harvey, and Renault (1996).

A regime switching model in continuous time is introduced where a variety of jumps are allowed, in addition to the diffusive component. The first contribution of the paper is the computation of the characteristic function of the process. This is derived in closed form, and is subsequently employed to create the likelihood function. Standard results of the option pricing literature can be employed in order to compute derivative prices. To this final end, the relationship between the physical and the risk adjusted probability measure is explored.

Intuitively, a Markov chain has the generic structure which is shared with the majority of stochastic volatility models, or other specifications with unobserved latent factors. The second contribution of the paper is to explore the ways that this relation can be explored in order to estimate, filter and carry out option pricing for such continuous state-space models. It is shown that virtually any stochastic volatility model model can be approximated arbitrarily well by a carefully chosen continuous time Markov chain. Thus, estimation, filtering and option pricing for stochastic volatility models can be carried out using the simple and numerically convenient Markov chain approximation.

The paper also identifies a number of interesting topics for future research. Although the family of the candidates for the approximating Markov chain is very

rich, and asymptotically they all share the same properties, the choice of the "optimal" chain for the finite samples deserves separate research. Turning to derivative pricing, issues arise when the impact of the risk aversion and the fact that the volatility process is latent are taken into account. Careful utility parameterizations are needed for these issues to be resolved. Finally, the proposed estimation approach demands to be formally compared to the existing algorithms for stochastic volatility jump diffusion estimation. Samples augmented with option prices will enjoy the full benefits of the presented approach which approximates not only the likelihood function, but also the relevant option prices.

References

Aït-Sahalia (2002). Telling from discrete data whether the underlying continuous-time model is a diffusion. *The Journal of Finance*, forthcoming.

Andersen, T. G., L. Benzoni, and J. Lund (1998). Estimating jump diffusions for equity returns. Technical report, Kellogg Graduate School of Management.

Bakshi, G. and D. Madan (2000). Spanning and derivative-security valuation. *Journal of Financial Economics 55*, 205–238.

Bates, D. S. (1995). Testing option pricing models. National Bureau of Economic Rresearch, Working Paper 5129.

Bates, D. S. (1997). Post-'87 crash fears in s&p500 futures options. National Bureau of Economic Rresearch, Working Paper 5894.

Benzoni, L. (1999). Pricing options under stochastic volatility: An econometric analysis. Kellogg Graduate School of Management, Northwestern University.

Bertoin, J. (1996). *Lévy Processes*. Cambridge Tracks in Mathematics. Cambridge, U.K.: Cambridge University Press.

Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen (1999). A new class of stochastic volatility models with jumps: Theory and estimation. Centre Interuniversitaire de Recherche en Analyse des Organisations, Working Paper 2002s–58.

Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen (2002). Alternative models for stock price dynamics. Centre Interuniversitaire de Recherche en Analyse des Organisations, Working Paper 99s–48.

Chernov, M. and E. Ghysels (2000). A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of option valuation. *Journal of Financial Economics 56*, 407–458.

Chiarella, C., S. Pasquali, and W. J. Runggaldier (2001). On filtering in Markovian term structure models (an approximation approach). *Advances in Applied Probability 33*(4), 794–809.

Cohrane, J. H. (1997). Where is the market going? Uncertain facts and novel theories. *Economic Perspectives 21*, 3–37.

Das, S. (1998). Poisson–Gaussian processes and the bond markets. Harvard Business School, Working Paper.

Dollard, J. D. and C. N. Friedman (1979). *Product Integration with Applications to Differential Equations*. Encyclopedia of Mathematics and its Applications. Reading, MA: Addison-Wesley.

Duffie, D., J. Pan, and K. Singleton (1999). Transform analysis and asset pricing for affine jump–diffusions. National Bureau of Economic Research, Working Paper 7105.

Dumas, B., J. Fleming, and R. Whaley (1998). Implied volatility functions: Empirical tests. *The Journal of Finance 6*, 2059–2106.

Dupuis, P. G. (2002). Markov chain approximations for stochastic nonlinear systems. IUTAMS Symposium on Nonlinear Stochastic Dynamics.

Dupuis, P. G. and H. J. Kushner (2001). *Numerical Methods for Stochastic Control Problems in Continuous Time*, Volume 24 of *Applications of Mathematics*. New York, N.Y.: Springer Verlag.

Eraker, B., M. Johannes, and N. Polson (2000). The impact of jumps in volatility and returns. Technical report.

Frey, R. and W. J. Runggaldier (2001). A nonlinear filtering approach to volatility estimation with a view towards high frequency data. *International Journal of Theoretical and Applied Finance 4*, 199–210.

Galland, R. A. and D. W. Nychka (1987). Seminonparametric maximum likelihood estimation. *Econometrica 55*, 363–390.

Gallant, A. R. and G. Tauchen (1997). EMM: A program for eficient method of the moments estimation. version 1.4 user's guide. Working Paper, University of North Carolina.

Ghysels, E., A. Harvey, and E. Renault (1996). Stochastic volatility. In G. Maddala and C. Rao (Eds.), *Handbook of Statistics, 14, Statistical Methods in Finance*. North Holland.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*, 357–384.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Harvey, A., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies 61*, 247–264.

Jackwerth, J. C. (1999). Option-implied risk-neutral distributions and implied binomial trees. *Journal of Derivatives 81*, 66–82.

Jackwerth, J. C. and M. Rubinstein (1996). Recovering probability distributions from options prices. *The Journal of Finance 51*, 1611–1631.

Jacquier, E., N. G. Polson, and P. G. Rossi (1995). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics 12*, 371–417.

Jarrow, R. A., D. Lando, and S. M. Turnbull (1997). A Markov model for the term structure of credit risk spreads. *Review of Financial Studies 10*(2), 481–523.

Jondeau, E. and M. Rockinger (1998). Estimating Gram-Charlier expansions under positivity constraints. Cahier de Recherche du Groupe HEC, Working Paper 653/1998.

Kendal, M. and A. Stuart (1977). *The Advanced Theory of Statistics* (4th ed.), Volume I. London, U.K.: Charles Griffin and Co.

Kooperberg, C. and C. J. Stone (1991). A study of logspline density estimation. *Computational Statistics and Data Analysis 12*, 327–347.

Kushner, H. J. (1977). *Probability Methods for Approximation in Stochastic Control and Elliptic Equations*. New York: Academic Press.

Kushner, H. J. (1990). Numerical methods for stochastic control problems in continuous time. *SIAM Journal of Control and Optimization 28*(5), 999–1048.

Kushner, H. J. and G. DiMasi (1978). Approximations of functionals and optimal control problems on jump diffusion processes. *Journal of Mathematical Analysis and Applications 63*, 772–800.

Laurent, J.-P. and D. P. J. Leisen (2000). Building a consistent pricing model from observed option prices. In *Collected Papers of the New York University Mathematical Finance Seminar*, Number 2. World Scientific.

Lo, A. W. and J. Wang (1995). Implementing option pricing models when asset returns are predictable. *The Journal of Finance 50*, 87–129.

Mathias, R. (1997). A chain rule for matrix functions and applications. mimeo: http://www.math.wm.edu/ mathias/.

Merton, R. (1976). Option pricing when the underlying stock returns are discontinuous. *Journal of Financial Economics 4*, 125–144.

Naik, V. (1993). Option valuation and hedging strategies with jumps in the volatility of asset returns. *The Journal of Finance 48*, 1969–1984.

Nelson, D. B. (1990). ARCH models as diffusion approximations. *Journal of Econometrics 45*, 7–39.

Nelson, D. B. and D. P. Foster (1994). Asymptotic filtering theory for univariate ARCH models. *Econometrica 62*, 1–41.

Pan, J. (1997). Stochastic volatility with reset at jumps. Permanent Working Paper.

Pan, J. (2002). The jump risk-premia implicit in options: Evidence from an integrated time series study. *Journal of Financial Economics 63*, 3–50.

Piccioni, M. (1987). Convergence of implicit discretization schemes for linear differential equations with application to filtering. In G. D. Prato and L. Tubaro (Eds.), *Stochastic PDE's and Applications, Lecture Notes in Mathematics 1236*. Springer-Verlag.

Scott, L. (1987). Option pricing when the variance changes randomly: Theory, estimation and an application. *Journal of Financial and Quantitative Analysis 22*, 419–438.

Singleton, K. (1998). Estimation of affine diffusion models based on the empirical characteristic function. University of Georgia, Working Paper.

Veronesi, P. (2001). Belief-dependent utilities, aversion to state-uncertainty and asset prices. University of Chicago, Manuscript.

Proofs

Proof of proposition 1

Let $\varepsilon > 0$, and built the ε–skeleton, $\mathbf{x}_n^{(\varepsilon)}$ of $\mathbf{x}(t)$. Consider a partition of the time interval $[0, t]$ into N subintervals of length ε, of the form $[n\varepsilon, n\varepsilon + \varepsilon]$, for $0 \leqslant n \leqslant M$. Since we are interested in the limiting behavior as $\varepsilon \downdownarrows 0$, without loss of generality one can make the assumption that $(M + 1)\varepsilon = t$. On the boundaries of these subintervals, the continuous time Markov chain (CTMC) and its skeleton coincide, that is to say $\mathbf{x}(n\varepsilon) = \mathbf{x}_n^{(\varepsilon)}$. Now, define the random variables $w_n^{(\varepsilon)}$, which are distributed with respect to their Laplace transforms given in (2) if the skeleton remains at the same state, and (3) if the state of the skeleton changes. Again since we are interested in the behavior as $\varepsilon \downdownarrows 0$, one can choose ε so small, such that during any subinterval $[n\varepsilon, n\varepsilon + \varepsilon]$ there is at most one state change and at most one jump. Then the discrete–time process $\sum_{n=1}^{M} w_n^{(\varepsilon)} = S^{(\varepsilon)}(M) \to S(t)$, as $\varepsilon \downdownarrows 0$ [or $M \to \infty$]. Consider a time period n, $0 \leqslant n \leqslant M$ for the ε–skeleton, and the corresponding time $u = n\varepsilon$ for the CTMC, where the two chains coincide. Denote the conditional characteristic function $\phi_{ji}^{S}(\theta, u) = \mathsf{E}\{e^{\varpi\theta S(u)} | \mathbf{x}(0) = \mathbf{e}_i, \mathbf{x}(u) = \mathbf{e}_j\}$. Define the event $\mathscr{E} = \{\mathbf{x}_0^{(\varepsilon)} = \mathbf{e}_j, \mathbf{x}_{n-1}^{(\varepsilon)} = \mathbf{e}_l, \mathbf{x}_n^{(\varepsilon)} = \mathbf{e}_k\}$, and observe that it can be rewritten as $\mathscr{E} = \{\mathbf{x}(0) = \mathbf{e}_j, \mathbf{x}(u - \varepsilon) = \mathbf{e}_l, \mathbf{x}(u) = \mathbf{e}_k\}$. Conditional on the event $\mathscr{F}(t) \otimes \mathscr{E}$, the random variables $S^{(\varepsilon)}(n-1)$ and $w_{n-1}^{(\varepsilon)}$ are mutually independent, and the characteristic function $\phi_{kl}^{w}(\theta, \varepsilon)$ will satisfy

$$\phi_{kl}^{w}(\theta, \varepsilon) = \begin{cases} 1 + \Psi_l(\theta)\varepsilon + o(\varepsilon), & \text{when } l = k \\ \Psi_{kl}(\theta), & \text{when } l \neq k \end{cases}. \tag{15}$$

Independency implies that $\mathsf{E}\{e^{\varpi\theta S(u)} | \mathscr{F}(t) \otimes \mathscr{E}\} = \phi_{lj}^{S}(\theta, u-\varepsilon)\phi_{kl}^{w}(\theta, \varepsilon)$, or by defining the n–period transition of the ε–skeleton, from state $\mathbf{e}_i$ to $\mathbf{e}_j$ to be given by $p_{ji}(u)$,

$$\mathsf{E}\{e^{\varpi\theta S(u)} | \mathbf{x}(0) = \mathbf{e}_j, \mathbf{x}(u) = \mathbf{e}_k\} = \frac{1}{p_{kj}(u)} \sum_{l} p_{lj}(u - \varepsilon)\phi_{lj}^{S}(\theta, u - \varepsilon)p_{kl}(\varepsilon)\phi_{kl}^{w}(\theta, \varepsilon).$$

Rewrite the above as

$$p_{kj}(u + \varepsilon)\phi_{kj}^{S}(\theta, u + \varepsilon) = \sum_{l} [p_{lj}(u)\phi_{lj}^{S}(\theta, u)][p_{kl}(\varepsilon)\phi_{kl}^{w}(\theta, \varepsilon)],$$

and denote $\tilde{\phi}_{ji}(\theta, u) = p_{ji}(u)\phi_{ji}^{S}(\theta, u)$. Observe that, as ε gets sufficiently small, one can approximate $p_{kl}(\varepsilon) = \delta(l - k, 0) + q_{kl}\varepsilon + o(\varepsilon)$. Hence the expectation will be written as

$$\frac{\tilde{\phi}_{kj}(\theta, u + \varepsilon) - \tilde{\phi}_{kj}(\theta, u)}{\varepsilon} = \sum_{l \neq k} \tilde{\phi}_{kl}(\theta, u) q_{kl} \Psi_{kl}(\theta) + \tilde{\phi}_{kj}(\theta, u)[q_{kk} + \Psi_k(\theta)].$$

Passing to the limit, $\varepsilon \Downarrow 0$, and collecting the elements $\tilde{\phi}_{kj}(\theta, u)$ in a matrix $\tilde{\mathbf{\Phi}}(\theta, u)$, yields

$$\frac{\partial}{\partial u}\tilde{\mathbf{\Phi}}(\theta, u) = \tilde{\mathbf{\Phi}}(\theta, u)\mathbf{B}(\theta), \tag{16}$$

where the matrix $\mathbf{B}(\theta)$ has elements of the form (15). The matrix differential equation resembles a system of $(N \times N)$ partial differential equations, with trivial boundary conditions $\tilde{\mathbf{\Phi}}(\theta, 0) = \mathbf{I}_N$, which are satisfied by the matrix exponential above. Bayes' rule implies the form of the weighted solution $\phi(\theta, \tau)$ given in Theorem 1. ▲

Proof of lemma 2

Under this assumption , the filtered probabilities $\xi(t|t)$ are the same under the physical and the risk adjusted measure. Suppose that the Radon-Nikodym derivative of the risk adjusted probabilities with respect to the physical ones is given by the set $\Xi_j{}_{j=1}^{N}$, which implies that $\bar{\xi}_j(t|t) = \Xi_j \xi_j(t|t)$. It is easy to verify that since this relationship has to hold for every physical distribution $\xi_j(t|t)_{j=1}^{N}$, the Radon-Nikodym derivative has to be equal to one. Equivalently the filtered physical probabilities are equal to the risk adjusted ones. ▲

Proof of lemma 3

Consider the random process $x_j(t)$, the j–th element of the Markov chain $\mathbf{x}(t)$. Fix $\delta > 0$, and denote with $\xi(0)$ the initial distribution of the states, with elements $\xi_j(0)$. With probabilities which are greater than $o(\delta)$ the quantity $\Delta x_j(0) = x_j(\delta) - x_j(0)$ can take the following values

$$\begin{aligned}
\Delta x_j(0) &= +1, \text{ in the event } x_k(0) = 1, x_j(\delta) = 1 \text{ for some } k \neq j \\
\Delta x_j(0) &= -1, \text{ in the event } x_j(0) = 1, x_k(\delta) = 1 \text{ for some } k \neq j \\
\Delta x_j(0) &= 0, \text{ in all other cases}
\end{aligned}$$

which imply that the conditional on $\mathscr{F}(0)$ expectation under risk neutrality is equal to

$$\bar{\mathsf{E}}_0\{\Delta x_j(0)\} = \sum_{k \neq j} \left[\xi_k(0)\bar{q}_{jk} - \xi_j(0)\bar{q}_{kj}\right]\delta + o(\delta)$$

Alternatively, the expectation can be written in terms of the marginal utility of wealth as

$$\begin{aligned}
\bar{\mathsf{E}}_0\{\Delta x_j(0)\} = \mathsf{E}_0\{\Delta x_j(0)(1 + \Delta\mathcal{M}(0))\} = \\
\sum_{k \neq j} \left[\xi_k(0)q_{jk}(1 + \Delta\mathcal{M}_{jk}(0)) - \xi_j(0)q_{kj}(1 + \Delta\mathcal{M}_{kj}(0))\right]\delta + o(\delta)
\end{aligned}$$

If one equates the above relationships and observes that they have to hold for all j and all initial distributions $\xi(0)$, one will reach the result of lemma 3 ▲

Proof of lemma 4

Consider the random process $Z(t)$. Fix $\delta > 0$ and consider the random variable $\Delta Z(0) = Z(\delta) - Z(0)$. With probabilities greater than $o(\delta)$ the quantity $\Delta Z(0)$ can follow the densities that have means

$$\mathsf{E}_0\{\Delta Z(0)\} = \mu_{jk} \text{ and } \bar{\mathsf{E}}_0\{\Delta Z(0)\} = \bar{\mu}_{jk},$$
$$\text{in the event } x_k(0) = 1, x_j(\delta) = 1 \text{ for all pairs } k, j \text{ with } j \neq k$$
$$\Delta Z_j(0) = 0, \text{ in all other cases}$$

The expectation of $\Delta Z(0)$ under the risk adjusted measure will therefore be

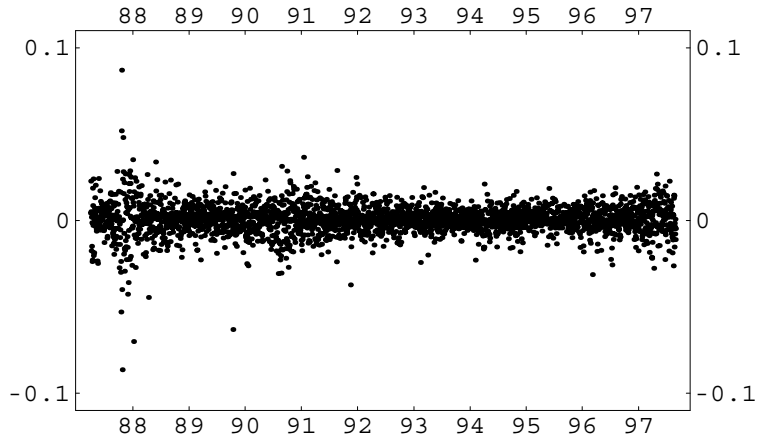
$$\bar{\mathsf{E}}\{\Delta Z(0)\} = \sum_k \sum_{j \neq k} \xi_j(0) \bar{q}_{jk} \bar{\mu}_{jk} \delta + o(\delta)$$

Alternatively, the expectation can be written using the marginal utility as

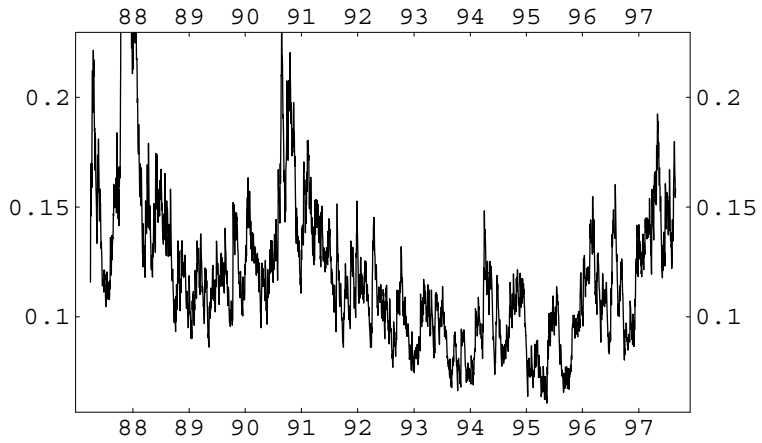
$$\bar{\mathsf{E}}_0\{\Delta Z(0)\} = \sum_k \sum_{j \neq k} \xi_j(0) q_{jk} \mathsf{E}_0\{\Delta Z_{jk}(0)(1 + \Delta \mathcal{M}_{jk}(0))\}\delta + o(\delta)$$

Substituting the risk adjusted transition probabilities, and recognizing that the equality has to hold for all distributions $\xi(0)$ gives the result of lemma 4 ▲

List of Figures

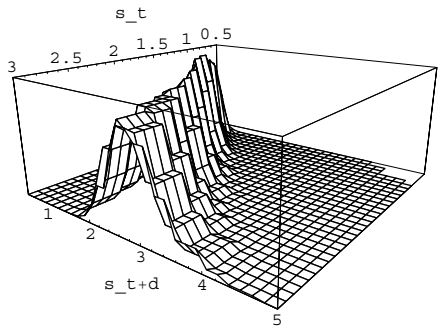


(a) Log-return series

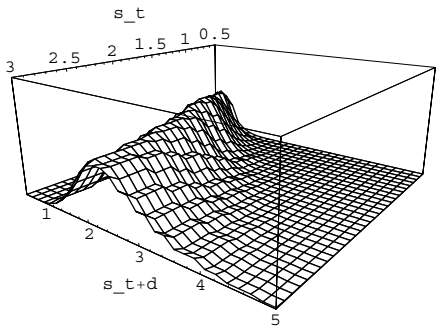


(b) Volatility series

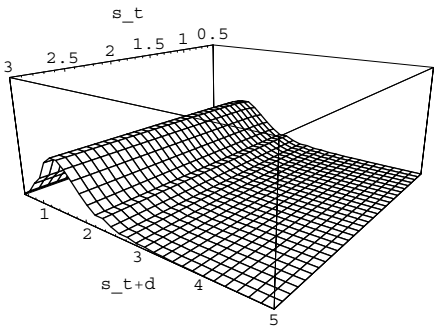
Figure 1: Whole sample series. (a) The log-return series of the SP500 index. Values around the '87 crash are omitted for scaling purposes. (b) The filtered volatility series calculated as the expectation of the Markov chain conditional on the information that prevailed at the time. The series is computed using the estimated parameters of the jump diffusion stochastic volatility model SVρJ, and includes the volatility contribution of the jump component. Values around the '87 crash are omitted for scaling purposes.



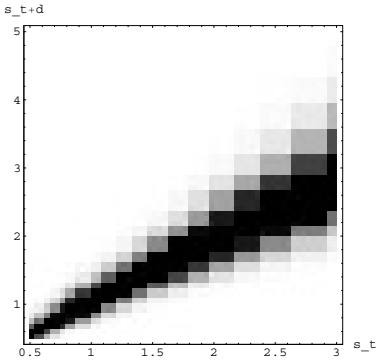
(a) Transition Kernel – 1 week



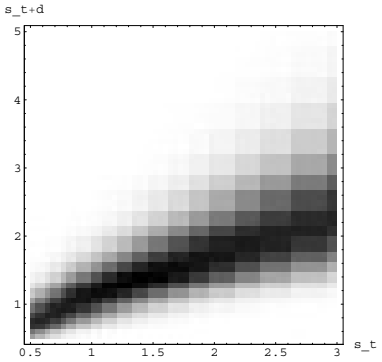
(b) Transition Kernel – 1 month



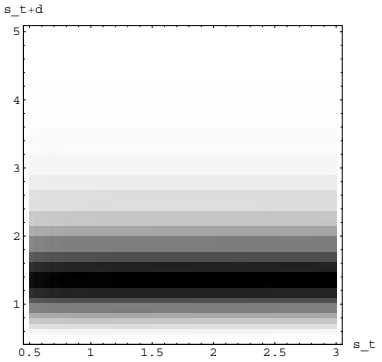
(c) Transition Kernel – 1 year



(d) Transition Kernel – 1 week

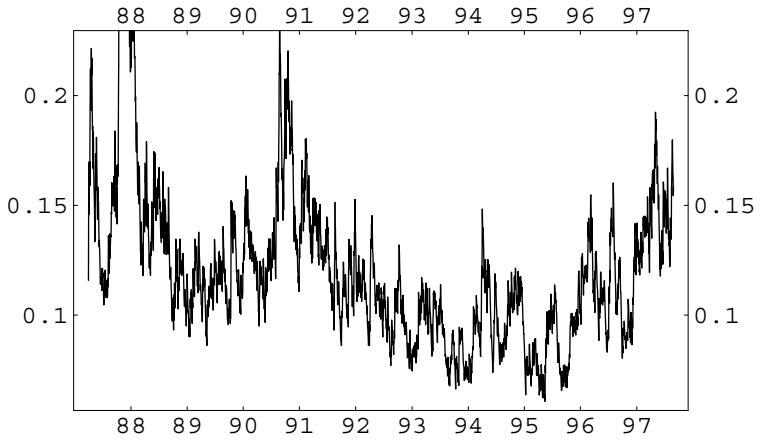


(e) Transition Kernel – 1 month

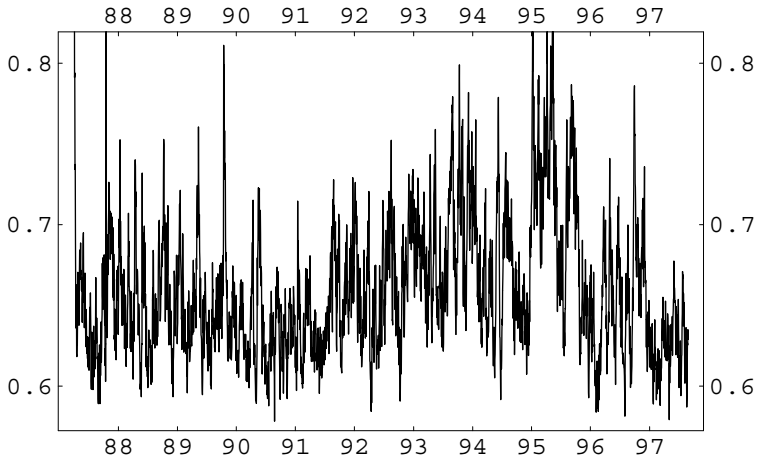


(f) Transition Kernel – 1 year

Figure 2: Transition kernels for the volatility diffusion in the jump diffusion stochastic volatility model SVρJ. As the forecasting horizon increases towards one year, it is apparent that the transition kernels loose their dependency on the starting volatility and converge to their ergodic distributions.

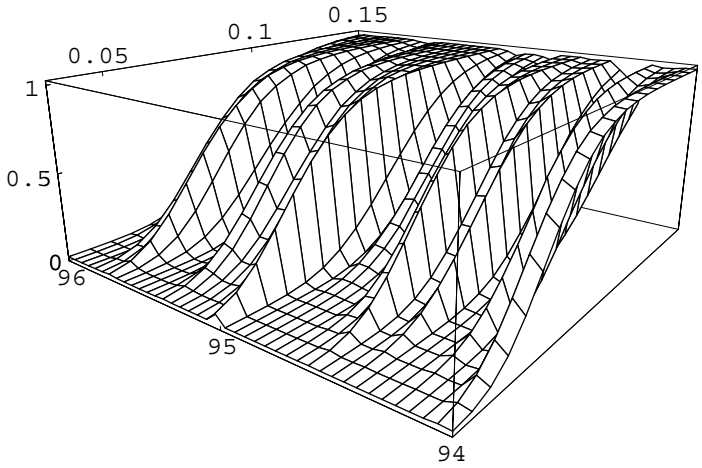


(a) Volatility series

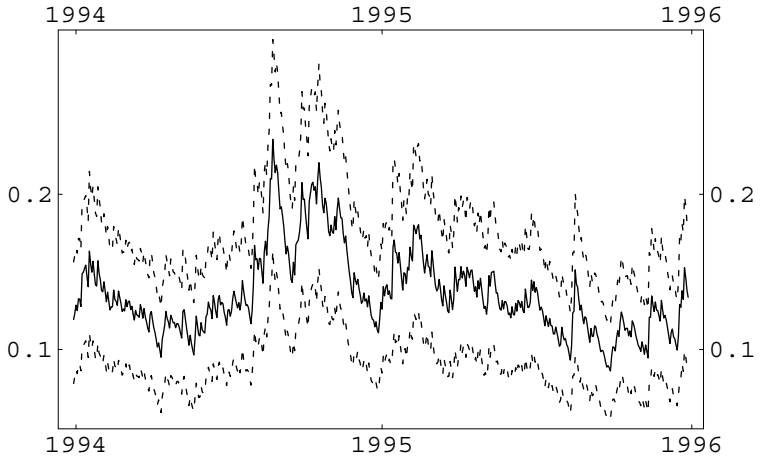


(b) Volatility relative spread

Figure 3: Filtered volatility series for the whole sample. (a) The filtered volatility series calculated as the expectation of the Markov chain conditional on the information that prevailed at the time; and (b) the relative spread. All series are computed using the estimated parameters of the jump diffusion stochastic volatility model SVρJ. All volatility figures do *not* include the jump component, and therefore do *not* represent the variance of the filtered series, which is given in figure 1.



(a) 94-96 Distribution



(b) 94-96 Spread

Figure 4: Filtered volatility series for the period 1994-1996. (a) The cumulative distribution of the volatility across time; (b) The filtered volatility together with the corresponding 5% and 95% bounds. All series are computed using the estimated parameters of the jump diffusion stochastic volatility model SVρJ. All volatility figures do *not* include the jump component, and therefore do *not* represent the variance of the filtered series, which is given in figure 1.

List of Tables

	$\mathscr{L}$	μ	θ	$\ln \overline{\sigma}^2$	ϕ	ρ	λ	$\mu_J(\%)$	$\sigma_J(\%)$
SV	9075.35	0.1783	8.4174	−4.3313	3.1335	$\equiv 0$	$\equiv 0$		
		(0.0059)	(0.2138)	(0.0865)	(0.1473)				
SVρ	9084.54	0.1916	5.2542	−4.3365	2.3577	−0.6147	$\equiv 0$		
		(0.0004)	(0.0110)	(0.0091)	(0.0050)	(0.0013)			
SVρJ	9103.85	0.1709	5.5084	−4.3006	2.0817	−0.6654	13.95	−4.858	2.246
		(0.0003)	(0.0110)	(0.0064)	(0.0041)	(0.0014)	(0.032)	(0.000)	(0.005)

Table 1: Estimation Results of stochastic volatility jump diffusions. The data set consists of daily returns on the SP500 index. Standard errors, constructed by numerically inverting the Hessian matrix, are reported in parantheses.

Model	$MSE^{\ddagger}$	a_0	a_1	a_2	a_3
A	0.7278	$0.21^{\ddagger}$	-0.0177		
		(0.05)	(0.0037)		
B	0.7298	$-1.03^{\dagger}$		$1.56^{\dagger}$	
		(0.26)		(0.39)	
C	0.6395	$0.00^{\ddagger}$			-0.1536
		(1.55)			(0.0078)
D	0.6396	$0.18^{\dagger}$	$-0.29\dagger$		1.1551
		(0.24)	(0.37)		(0.0080)

Table 2: Estimates of the nested models: $\Delta \mathsf{E}_{\xi}(t) = a_0 + a_1 \mathsf{E}_{\xi}(t) + a_2 \mathsf{RSP}_{\xi}(t) + a_3 \Delta \mathsf{RSP}_{\xi}(t)$, where $\mathsf{E}_{\xi}(t)$ is the estimated filtered volatility and $\mathsf{RSD}_{\xi}(t)$ is the relative spread of the estimator. The parameter estimates are for the full "SVρJ" specification. Standard errors are reported in parentheses. † denotes $\times 10^{-2}$; ‡ denotes $\times 10^{-4}$.

This working paper has been produced by
the Department of Economics at
Queen Mary, University of London

Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096 or Fax: +44 (0)20 8983 3580
Email: j.conner@qmul.ac.uk
Website: www.econ.qmul.ac.uk/papers/wp.htm