NBER WORKING PAPER SERIES

TELLING FROM DISCRETE DATA WHETHER THE UNDERLYING CONTINUOUS-TIME
MODEL IS A DIFFUSION

Yacine Ait-Sahalia

Telling from Discrete Data Whether the Underlying Continuous-Time
Model is a Diffusion
Yacine Ait-Sahalia
NBER Working Paper No. 8504
October 2001
JEL No. G12, C22, C52

ABSTRACT

Asset returns have traditionally been modeled in the literature as following continuous-time
Markov processes, and in many cases diffusions. Can *discretely sampled* financial rate data help us decide
which *continuous-time* models are sensible? Diffusion processes are characterized by the continuity of
their sample paths. This cannot be verified from the discrete sample path: by nature, even if the
underlying sample path were continuous, the discretely sampled data will always appear as a sequence
of discrete jumps. Instead, this paper relies on a characterization of the transition density of the discrete
data to determine whether the discontinuities observed in the discrete data are the result of the
discreteness of sampling, or rather evidence of genuine jump dynamics for the underlying continuous-
time process. I then focus on the implications of this approach for option pricing models.

Yacine Ait-Sahalia
Department of Economics / Bendheim Center for Finance
Princeton University
Princeton, NJ 08540-5296
and NBER
609-258-4015
yacine@princeton.edu

In many instances in financial econometrics, we make inference about a postulated continuous-time model on the basis of discretely sampled observations. Among potential continuous-time models, most specifications adopted have been diffusions, although the literature is more and more frequently allowing for jumps (see Ahn and Thompson (1988), Bates (1991), Das and Foresi (1996), Duffie, Pan and Singleton (2000), Aït-Sahalia, Wang and Yared (2001) among others).

A diffusion process is a Markov process with continuous sample paths. Suppose we observe the process every Δ units of time, with Δ not necessarily small. Presented with such a discrete subsample of the continuous-time path, can we tell whether the underlying model that gave rise to the data was a diffusion, or should jumps be allowed into the model? Intuition suggests that the answer should be no. After all, the discrete data are purely discontinuous even if the continuous-time sample is not so, faced with two discontinuous samples, how could we ever rule out that one came from a diffusion but not the other? It turns out that this question is not as hopeless as it first sounds. A finer look reveals that there are different degrees of discontinuity in the discrete observations, some compatible with the continuity of the underlying sample path, some not.

The approach I use relies on identifying a necessary and sufficient restriction on the transition densities of diffusions, at the sampling interval of the observed data. The theory is based on Karlin and McGregor (1959b)'s notion of coincidence probabilities combined with crossing arguments. The total positivity restriction characterizes the continuity of the unobservable complete sample path. It must be satisfied by the transitions densities of a diffusion and is valid for every sampling interval including long ones.

In a nutshell, the argument is based on the fact that if a diffusion on the real line starts below another diffusion, it cannot finish above the second one without their sample paths having crossed at least once.[1] Since the discrete data reveal the transition density at whatever sampling interval is available, one can actually discriminate between diffusions and non-diffusion Markov processes on the basis of their discrete subsamples.

---

[1] An alternative property is that the eigenvalues of the infinitesimal generator of the diffusion are all real and nonnegative (see Mandl (1968)). This property has been exploited by Florens, Renault and Touzi (1998). As a practical matter, however, eigenvalues and eigenfunctions of generators can be difficult to calculate. The first eigenvalue and eigenfunctions can be determined only in certain special cases (see e.g., Aït-Sahalia (1996a) and Hansen, Scheinkman and Touzi (1998)).

An essential property of the characterization is that if the transitions over an interval of length Δ satisfy the inequality characterization, then longer transitions will satisfy it as well. That is, we only need to focus on a single criterion, rather than attempt to verify a large number of conditions, to determine whether the process is a diffusion. Furthermore, the criterion can determine whether a discrete transition function is compatible with *some* diffusion without requiring that a list of potential candidate diffusions be exhibited.

To provide some intuition, I give the corresponding version of the criterion for discrete-state models, both for continuous-time Markov chains and discrete-time trees. In a discrete-state world, by definition, all the state changes are jumps and the appropriate notion of continuity distinguishes between small or continuous jumps, which are those from one state to an immediately adjacent one, from large or discontinuous jumps, which are those from one state to a nonadjacent one. One consequence for trees in option pricing is that binomial and trinomial trees with branches continuously spaced are inappropriate as approximations to jump-diffusions, despite their common use in that context.

Finally, I employ the criterion function to determine whether the risk-neutral transition density of the S&P 500 implied by observed option prices is compatible with an underlying continuous-time diffusion for the index, or whether jumps should be included. The latter happens to be the case empirically. I examine the consequences of this finding for the implied diffusion, implied tree and Edgeworth expansion approaches that are widely used in practice to price and hedge equity derivatives. Another possible empirical motivation, which is not pursued here, would be to substantiate, or invalidate, the approach of modeling the dynamics of the short term interest rate, or other factors taken individually, as continuous-time diffusions: should they be diffusions, or something else within the Markov class?

The paper is organized as follows. Section I examines the implications of the continuity of the sample paths for the discrete data and obtains a necessary and sufficient characterization of the transition density of diffusions. Section II interprets this criterion in terms of discrete-state processes. Section III gives two examples and two consequences. Section IV focuses on the dynamics implicit in S&P 500 option prices. Section V summarizes the results and concludes.

I. Implications of the Continuity of the Sample Paths for the Discrete Data

A. Sample Path – Level Characterization of Continuity

I start with some mathematical preliminaries and by establishing the notation used throughout the paper. Let us assume that the process $\{r_t, t \geq 0\}$ is Markovian. That is, the continuous-time process $\{r_t, t \geq 0\}$ is defined on a probability space $(\Omega, \Im, P)$, takes values in $D \subseteq \mathbb{R}$ and, with $F_t \equiv \sigma(r_s : s \leq t)$, assume that

$$P(r_{t+s} \in I \mid F_t) = P(r_{t+s} \in I \mid r_t) \tag{1}$$

for all t, s ≥ 0 and open interval I in D. This by itself can be an interesting hypothesis to examine (see the companion paper Aït-Sahalia (2000)), but will be maintained throughout this paper. We are then interested in learning, from discrete observations $\{r_0, r_\Delta, r_{2\Delta}, \ldots, r_{n\Delta}\}$, whether it belongs to the smaller class of diffusion processes. Note that I restrict attention to univariate processes, so the state space is an interval D on the real line.

Let $p(s, y \mid t, x)$ be the conditional density of $r_s = y$ given $r_t = x$. This is the transition function of the process. To be able to infer the transition function from a time series of observations on r, we must assume that the joint densities of the process are time-homogenous. That is, each pair of observations $(r_{t+\Delta}, r_t)$ at each date t are drawn from the same joint density, which I denote as $p(\Delta, y, x)$ and is independent of t. This assumption lets us in effect transform what would have been a single data point, the observed path, into repeated observations on the pair $(r_{t+\Delta}, r_t)$ drawn from a common distribution (see Figure 1).

Then let $p(\Delta, y \mid x)$ denote the transition function of the process over a time interval of length Δ, i.e., the conditional density of $r_{t+\Delta} = y$ given $r_t = x$. Without time-homogeneity, the transition function would be $p(t + \Delta, y \mid t, x)$, i.e., a function of t and Δ separately. This does not mean that we are only considering stationary processes. Time-homogeneity is necessary for the stationarity of the process, i.e., all finite dimensional distributions being identical no matter where taken in time, but not sufficient. For instance, a Brownian motion is time-homogenous but of course not stationary.

I further assume that the sample paths of r are right-continuous, so the finite-dimensional distributions determine the probabilities of all events, and that r is a strong Markov process. That is, r restarts at the first passage of a given point and retains its transition densities afterwards. More precisely, recall that F_s is the σ-algebra generated by $\{r_t, 0 \leq t \leq s\}$. A stopping time of the

process $\{r_t, t \geq 0\}$ is a random variable τ such as the events $\{\tau \leq t\}$ belong to F_t for all t≥0. Then $\{r_t, t \geq \tau\}$ is a Markov process independent of F_t and has the same transition densities as before τ. A sufficient condition on the transition density p of the process which guarantees that it is a strong Markov process is Feller's property: for every bounded continuous real function f, and for every t>0, $x \mapsto \int_D f(y)p(t, y \mid x)dy$ defines a continuous function on D (see Ray (1956) and Friedman (1975), Theorem 2.2.4).

Note also that the transition densities of a Markov process must satisfy the Chapman-Kolmogorov equation

$$p(2\Delta, y \mid x) = \int_{\underline{r}}^{\overline{r}} p(\Delta, y \mid z)p(\Delta, z \mid x)dz \tag{2}$$

for every x in D, Y⊂D, and Δ>0. This fact will be handy later.

Then a diffusion is a process satisfying the assumptions above and whose sample paths are continuous everywhere, except possibly for jumps from the boundaries of the state space D.

B. Transition – Level Characterization of Continuity

Given that the full path is not observable, the first step in the approach will be to move away from the sample path characterization of diffusions that was just given, and examine what can be said about their transitions. Since the process is Markovian, and its transitions are time-homogenous, the information contained in the discrete data can be summarized through the transition function p(Δ,y|x). The following example describes what an easy solution to the problem would be, if it were available. Suppose that we find, using any available data analysis technique, that the discrete data can be represented accurately by the conditional density

$$r_{t+\Delta} \mid r_t \sim N\left(\gamma_0 + \gamma_1 r_t, \delta_0^2\right) \tag{3}$$

i.e., a Gaussian transition density with affine mean and constant variance.

Can we then construct a continuous-time diffusion which, based on the empirical evidence (3), could have generated the data as a discrete sample off its continuous sample path? The answer is yes. Consider the diffusion

$$dr_t = \kappa\left(\alpha - r_t\right)dt + \sigma dZ_t \tag{4}$$

for which p(Δ,y|x) is Gaussian with $E\left[r_{t+\Delta} \mid r_t\right] = r_t + (\alpha - r_t)\exp[-\kappa\Delta]$ and $V\left[r_{t+\Delta} \mid r_t\right] = \sigma^2(1 - \exp[-2\kappa\Delta])/(2\kappa)$. Now set $\kappa = -\text{Ln}[1 - \gamma_1]/\Delta$, $\alpha = \gamma_0/(1 - \gamma_1)$ and $\sigma^2 = -2\delta_0^2\text{Ln}[1 - \gamma_1]/[(1 - \gamma_1^2)\Delta]$: the continuous-time diffusion is fully determined. Based on the available data, we cannot rule out (4) as a continuous-time model that could have generated the discrete data.

Unfortunately, such explicit calculations are impossible to conduct in most cases, since, in general, one cannot compute in closed-form the transition function p(Δ,y|x) implied by a particular diffusion model, or vice versa, although very accurate closed-form approximations can be formed, see Aït-Sahalia (1997, 1999). On the other hand, this explicit calculation, whenever available, provides a constructive answer to the problem in that not only do we get to answer that a diffusion could have generated the discrete data, but we also get to identify that diffusion.

Alternatively, if we knew what specific diffusion to look for, i.e., had an idea as to what functions μ and σ^2 to use in $dr_t = \mu(r_t)dt + \sigma(r_t)dZ_t$, then we could place a restriction on the discrete transitions of the process for any Δ, even without knowing in closed-form what its transition density is. As shown in Aït-Sahalia (1996b), time-homogeneity of the transition density can be exploited by re-writing the forward and backward Fokker-Planck-Kolmogorov equations: note that by stationarity $p(s, y \mid t, x) = p(s - t, y \mid 0, x) \equiv p(s - t, y \mid x)$ for any $s > t > 0$ and therefore $\partial p/\partial s = -\partial p/\partial t$. We can then eliminate derivatives of p with respect to $\Delta = $ s-t, i.e., the left hand side terms in:

$$\begin{cases} \dfrac{\partial p(\Delta, y \mid x)}{\partial \Delta} = -\dfrac{\partial}{\partial y}\big(\mu(y)p(\Delta, y \mid x)\big) + \dfrac{1}{2}\dfrac{\partial^2}{\partial y^2}\big(\sigma^2(y)p(\Delta, y \mid x)\big) \\ \dfrac{\partial p(\Delta, y \mid x)}{\partial \Delta} = \mu(x)\dfrac{\partial}{\partial x}\big(p(\Delta, y \mid x)\big) + \dfrac{1}{2}\sigma^2(x)\dfrac{\partial^2}{\partial x^2}\big(p(\Delta, y \mid x)\big) \end{cases} \quad (5)$$

for all x and y in the interior of D since the two right hand sides must be equal.[2] So the terms that could not be estimated with discrete data, i.e., the derivatives $\partial p/\partial \Delta$, are now gone. But if we wish to test whether p(Δ,y|x) could possibly be the discrete transition function from a diffusion, (5) is of no help, since obviously we cannot examine every possible pair of functions (μ,σ^2), unless by sheer luck we happened to stumble upon the right choice for (μ,σ^2).

---

[2]The rationale for eliminating the term $\partial p/\partial \Delta$ is that it this term cannot be estimated given data sampled at a fixed interval Δ. We can only estimate p(Δ|y,x), p(2Δ|y,x), etc., and their derivatives with respect to the starting and ending state levels x and y.

C. *Implications of Continuity for Small Time Transitions*

So, what can be said about the transition densities of a diffusion without pre-committing to specific choices of (μ, σ^2)? In other words, does the fact that the underlying model is a diffusion –*any* diffusion-- imply that its transition densities will have distinguishing features? I will maintain that the transition functions of the process preserve the probability, in the sense that

$$\lim_{\Delta \to 0} \quad \int_{|y-x| \leq \varepsilon} p(\Delta, y \mid x) dy = 1 \tag{6}$$

for any fixed $\varepsilon > 0$ and x in the interior of D. For technical reasons, let me also assume throughout that the convergence in equation (6) occurs at least at some unrestricted polynomial rate.[3]

The natural first approach is to examine whether the small time characterizations of continuity, i.e., those valid in the limit where Δ goes to zero, extend to any discrete time interval. For instance, it is known that the sample paths of r are almost surely continuous functions of t if and only if, for every compact interval $I \subseteq D$ and $\varepsilon > 0$, Lindeberg's condition

$$\lim_{\Delta \to 0} \quad \frac{1}{\Delta} \int_{|y-x| > \varepsilon} p(\Delta, y \mid x) dy = 0 \tag{7}$$

is satisfied uniformly in x on D (see Ray (1956)). This condition maps out the continuity of the sample path into a bound on the size of the probability of leaving a given neighborhood in the amount of time Δ; intuitively, this probability must be small as Δ goes to 0 if the sample paths are to remain continuous. Condition (7) says *how small* this probability must be as Δ gets smaller: the answer is $o(\Delta)$, that is, negligible compared to Δ. Condition (6) only says that it is of order $o(\Delta^\alpha)$ for some $\alpha > 0$. Condition (7) says that α must be greater than or equal to one.

In terms of deciding whether the discrete data came from a diffusion, condition (7) represents a step forward compared to the notion of continuity of the sample path. However it is also clear that this condition still cannot be used as a basis of a test for diffusions. Indeed, unless we are presented with ultra-high frequency data, we do not have the necessary data to examine

---

[3] This means that, as Δ goes to zero, the difference between the integral on the right hand side of equation (6) and its limit one is of order $o(\Delta^\alpha)$, for some $\alpha > 0$.

the behavior of the transition densities as the sampling interval Δ goes to zero.[4] Condition (7) does not restrict the observable transition function $p(\Delta, y \mid x)$ for the fixed sampling interval of the dataset.[5] An alternative is to focus on a sufficient characterization for the continuity of sample paths such as Kolmogorov's criterion: if there exist three constants β, γ, C >0 such that as Δ goes to 0

$$E\left[\left|r_{t+\Delta} - r_t\right|^{\beta}\right] \leq C\Delta^{1+\gamma} \tag{8}$$

then the process r (has a version which almost-surely) has continuous sample paths (see e.g., Friedman (1975), Theorem 1.2.2). For the same reason, this criterion can not be used either to form a test: it only restricts the transitions of the process over infinitesimal instants[6] and does not extend to longer time intervals.

How about checking whether the density is determined by its conditional mean and variance? The Markov process r is entirely determined by the two functions μ and σ^2 defined by the limits

$$\begin{cases} \lim_{\Delta \to 0} \dfrac{1}{\Delta} \int_{|y-x|<\varepsilon} (y-x) p(\Delta, y \mid x) dy = \mu(x) \\ \lim_{\Delta \to 0} \dfrac{1}{\Delta} \int_{|y-x|<\varepsilon} (y-x)^2 p(\Delta, y \mid x) dy = \sigma^2(x) \end{cases} \tag{9}$$

uniformly for x in the interior of D (see e.g., Feller (1971), Section X.4). But this is equally unhelpful, as it is only true in the limit where Δ goes to 0. If all discrete transitions were characterized by the first two moments, then they would all be Gaussian.

---

[4] Testing whether the underlying data generating process is a diffusion might not be a sensible thing to do with ultra-high frequency data given market microstructure noise (such as bid-ask bounces). This might be less of an issue with decimalization.

[5] Since the Markov property is assumed, we can imply the shorter transitions p(Δ/m,y|x), m$\geq$2, from the longer ones, p(Δ,y|x) by solving equation (2), i.e., solving $p(\Delta, y \mid x) = \int_{\underline{r}}^{\overline{r}} p(\Delta/2, y \mid z) p(\Delta/2, z \mid x) dz$ for the function p(Δ/2,y|x). To base a test on the behavior of p(Δ/m,y|x) as m$\to\infty$ would necessitate that this difficult numerical task be repeated a number of times.

[6] For alternative characterizations, all sharing a local character in Δ, see Gikhman and Skorohod (1969), section IV.5.

D. *Transition – Level Characterization of Continuity for Any Sampling Interval*

In light of the above discussion, we would like a characterization of diffusions that satisfies the following criteria: (D1) the characterization must not rely on observing very fine transitions in time of the process; (D2) unlike the approaches suggested just above, it must not require that either we already know what candidate drift and diffusion functions (μ, σ^2) to use or that the discrete transition happen to match those of the small set of diffusions for which a closed-form solution is available; (D3) it should be based on a necessary and sufficient characterization of diffusions; (D4) and require that a single property be checked, for the particular sampling interval Δ of the available data; (D5) but nevertheless be sufficient to ensure that the property is satisfied for the longer observable time intervals. By that, I mean that if we observe the process at say, the daily frequency, and verify that the property holds at that frequency, then it should automatically be the case that the property is also satisfied at the lower observable frequencies of one observation every two days, or one every three days, etc.

So, what works? The approach that I propose to use to discriminate between diffusions and non-diffusions on the basis of discrete-time information relies on total positivity of order 2 property of the transition function of a diffusion (see Karlin and McGregor (1959b)). This approach leads to identifying and then verifying the necessary and sufficient condition that the function $p(\Delta, y|x)$ must satisfy if it were to correspond to the discrete transition function of a continuous-time diffusion, without ever requiring that we identify in closed-form the diffusion. Unless the transition function $p(\Delta, y|x)$ fails to satisfy that condition, after accounting for the sampling noise of the estimator if any, the hypothesis that the discrete observations came from a diffusion will not be rejected.

What follows is a heuristic approach that delivers the main result of this theory. Consider two processes $\{r_t, t \geq 0\}$ and $\{\tilde{r}_t, t \geq 0\}$ on $\mathbb{R}$, having the same transition probability densities p, but otherwise independent. Suppose that $r_t = x$, while $\tilde{r}_t = \tilde{x}$ with $x < \tilde{x}$. The essential consequence of the continuity of sample paths is that, at any future date t+Δ, r cannot be above $\tilde{r}$ without their sample paths having crossed at least once between t and t+Δ. Consider two potential values at t+Δ, $y < \tilde{y}$, and two sets Y and $\tilde{Y}$ such that all values in Y are smaller that those in $\tilde{Y}$. The probability that $r_{t+\Delta} \in Y$ and $\tilde{r}_{t+\Delta} \in \tilde{Y}$, without their sample paths having ever crossed between t and t+Δ is (see Figure 2):

$$\Pr\left(r_{t+\Delta} \in Y, \tilde{r}_{t+\Delta} \in \tilde{Y}, \left\{\forall \tau \in [t, t+\Delta], r_\tau \neq \tilde{r}_\tau\right\} \middle| r_t = x, \tilde{r}_t = \tilde{x}\right)$$
$$= \Pr\left(r_{t+\Delta} \in Y, \tilde{r}_{t+\Delta} \in \tilde{Y} \middle| r_t = x, \tilde{r}_t = \tilde{x}\right) \qquad (10)$$
$$- \Pr\left(r_{t+\Delta} \in Y, \tilde{r}_{t+\Delta} \in \tilde{Y}, \left\{\exists \tau \in [t, t+\Delta] / r_\tau = \tilde{r}_\tau\right\} \middle| r_t = x, \tilde{r}_t = \tilde{x}\right)$$

To evaluate the second term on the right-hand side, let τ be the first coincidence time between t and t+Δ, i.e., the smallest τ such that $r_\tau = \tilde{r}_\tau$ (see Figure 3). Of course, we do not actually observe the time τ, since we do not see what happens between t and t+Δ; we simply know that τ exists as part of the event to be evaluated, and that τ is a stopping time by the strong Markov property discussed above.

By the reflection principle and the communality of distributions, after time τ, we can no longer tell which is r and which is $\tilde{r}$. Therefore we can interchange them as

$$\Pr\left(r_{t+\Delta} \in Y, \tilde{r}_{t+\Delta} \in \tilde{Y}, \left\{\exists \tau \in [t, t+\Delta] / r_\tau = \tilde{r}_\tau\right\} \middle| r_t = x, \tilde{r}_t = \tilde{x}\right)$$
$$= \Pr\left(\tilde{r}_{t+\Delta} \in Y, r_{t+\Delta} \in \tilde{Y}, \left\{\exists \tau \in [t, t+\Delta] / r_\tau = \tilde{r}_\tau\right\} \middle| r_t = x, \tilde{r}_t = \tilde{x}\right) \qquad (11)$$
$$= \Pr\left(\tilde{r}_{t+\Delta} \in Y, r_{t+\Delta} \in \tilde{Y} \middle| r_t = x, \tilde{r}_t = \tilde{x}\right)$$

where the last equality follows since the sample paths r and $\tilde{r}$ must have crossed between t and t+Δ, for r, having started below $\tilde{r}$ at t, to finish above it at t+Δ. This fact intrinsically characterizes the continuity of the sample paths. For instance, if the process can have jumps, r and $\tilde{r}$ may reverse order without ever crossing (see Figure 4).

Incidentally, Figures 3 and 4 illustrate why the argument is inherently univariate. If the process can evolve in three dimensions as opposed to being restricted to the plane, then the two replications r and $\tilde{r}$ can interchange order without jumping, and without ever crossing as they do in Figure 3. Just imagine in Figure 4 that, instead of representing a jump on the plane, the dotted part of the path of $\tilde{r}$ represents a bridge that goes continuously over, or under, the path of r. In that situation, the two paths of r and $\tilde{r}$ remain continuous throughout, yet the two processes have reversed order without crossing.

Back now to the univariate case, by independence of r and $\tilde{r}$, we have

$$\begin{cases} \Pr\left(r_{t+\Delta} \in Y, \tilde{r}_{t+\Delta} \in \tilde{Y} \middle| r_t = x, \tilde{r}_t = \tilde{x}\right) = P(\Delta, Y \mid x) P(\Delta, \tilde{Y} \mid \tilde{x}) \\ \Pr\left(\tilde{r}_{t+\Delta} \in Y, r_{t+\Delta} \in \tilde{Y} \middle| r_t = x, \tilde{r}_t = \tilde{x}\right) = P(\Delta, Y \mid \tilde{x}) P(\Delta, \tilde{Y} \mid x) \end{cases} \qquad (12)$$

where $P(\Delta, Y \mid x) \equiv \int_{y \in Y} p(\Delta, y \mid x) dx$ denotes the common cumulative distribution function of the two processes.

Consequently the probability that $r_{t+\Delta} \in Y$ and $\tilde{r}_{t+\Delta} \in \tilde{Y}$, without their sample paths having ever crossed between t and t+Δ, is

$$P(\Delta, Y \mid x) P(\Delta, \tilde{Y} \mid \tilde{x}) - P(\Delta, Y \mid \tilde{x}) P(\Delta, \tilde{Y} \mid x) > 0. \qquad (13)$$

The inequality follows from the fact that the probability of any possible event is positive. For the inequality to be strict, assume that every transition is possible in the sense that $P(\Delta, Y \mid x) > 0$ for every x in D, Y⊂D, and Δ>0, i.e., the process is strict. Otherwise, replace strictly greater by greater or equal to in inequality (13). Intuitively, this inequality states that the probability that the relative ranking of r and $\tilde{r}$ remain unchanged, between t and t+Δ, is greater than the probability that their ranking be reversed. If the process is a diffusion, then inequality (13) must be satisfied for every $x < \tilde{x}$ and $Y < \tilde{Y}$ in the state space.

Assuming that the density function $p(\Delta, y \mid x)$ is continuous in y for each x, it follows from inequality (13) that the transition function of any diffusion process must obey the inequality

$$\delta(\Delta, y, \tilde{y} \mid x, \tilde{x}) \equiv p(\Delta, y \mid x) p(\Delta, \tilde{y} \mid \tilde{x}) - p(\Delta, y \mid \tilde{x}) p(\Delta, \tilde{y} \mid x) > 0 \qquad (14)$$

for any $x < \tilde{x}$ and $y < \tilde{y}$ in D.

E. Properties of the Transition – Level Characterization of Continuity

I now verify that the total positivity characterization (14) of diffusions satisfies the criteria that I had set earlier to be the basis of a discriminating criterion for discretely sampled diffusions. Firstly, inequality (14) is valid for any Δ, not just infinitesimally small ones, so (D1) is satisfied. Secondly, (D2) is obviously verified as well: unlike condition (5), inequality (14) makes no reference to the unknown μ and σ^2 functions of the diffusion process. I now check that this inequality is not only necessary, but also sufficient to characterize a diffusion process:

Proposition 1: If its transition densities satisfy the inequality (14) for every Δ>0, then the process is a diffusion.

Therefore, if inequality (14) holds for all Δ>0, then the process r has (almost surely) continuous sample paths. Hence the characterization of diffusions (14) satisfies (D3). Determining whether the discrete data could have come from a diffusion will be based on

checking the property (14), or its equivalent forms (13) or (15) below, for the sampling interval Δ corresponding to that of the data, thereby satisfying (D4). All that remains to be proved is that (D5) is satisfied, i.e.:

Proposition 2: If inequality (14) holds for the sampling interval Δ, then it must hold for longer observable time intervals as well.

Thus all five requirements are satisfied. Finally, condition (14) can be expressed in even simpler terms when the transition function is smooth:

Proposition 3: Assume that $p(\Delta, y|x)$ is strictly positive and twice-continuously differentiable on $D \times D$. Then (14) is equivalent to:

$$\frac{\partial^2}{\partial x \partial y} Ln(\, p(\, \Delta, y\,/\,x\,)) > 0 \quad \text{for all } \Delta > 0 \text{ and } (x,y) \in D \times D. \tag{15}$$

In the rest of the paper, I will refer to inequality (15) as the "diffusion criterion." Given a transition function $p(\Delta, y|x)$, satisfying (15) is equivalent to the proposition that the underlying continuous-time model that gave rise to those discrete transitions was a diffusion. For a given Δ, finding a pair of (x,y) where the criterion fails is sufficient (absent sampling noise) to reject the hypothesis that the underlying model could have been a diffusion.

One final remark. We know that if X is a diffusion, then any deterministic, twice-continuously differentiable and strictly monotonic function of X will also be a diffusion.[7] It would be desirable for the diffusion criterion to also satisfy this invariance property. This is indeed the case. Namely, we have that:

Proposition 4: The criterion (15) is invariant with respect to Itô transformations of the process.

---

[7] Such transformations are commonly used in finance, for instance to go from an arithmetic Brownian motion to a geometric Brownian motion, from a Cox-Ingersoll-Ross square-root process to a Bessel process, from a CEV process (with geometric mean) to a Bessel process, etc. Indeed, most closed-form solutions we rely on are obtained through such a transformation.

II. Interpretation: Discrete-State Processes

A. Continuous-Time, Discrete-State Markov Chains

To help interpret condition (14), or equivalently (15), and understand its implications for trees in derivative pricing, consider a continuous-time, stationary, Markov chain that can only take countable discrete values, say, {...,-1,0,1,...}. When does such a process have continuous sample paths? Obviously, the notion of continuity of a sample path depends on the state space: in $\mathbb{R}$, this is the usual definition of a continuous function. More generally, by continuity one means continuity with respect to the order topology of the state space of the process. In a discrete state space, the appropriate notion of continuity of the chain's sample paths is the following intuitive one: it constraints the chain to never jump by more than one state at a time, either up or down. It turns out that the restriction on the chain's transition probabilities analogous to (14) characterizes precisely this form of continuity.

Specifically, assume that the Markov chain is right-continuous and let $J_0, J_1,...$ be the jump times of the chain $X = \{X_t\}_{t \geq 0}$, defined by

$$J_0 = 0, \quad J_{n+1} = \inf\left\{ t \geq J_n / X_t \neq X_{J_n} \right\} \tag{16}$$

for n=0,1,..., with the convention that $\inf\{\emptyset\} = \infty$ (see Figure 5). Consider now the restriction for the transition matrix of the Markov chain that is analogous to condition (14) for the conditional density of a diffusion, namely

$$
\begin{aligned}
\delta\left(\Delta, j, \tilde{j} \mid i, \tilde{i}\right) \quad &\equiv \quad \Pr\left(X_{t+\Delta} = \tilde{j} \mid X_t = \tilde{i}\right) \Pr\left(X_{t+\Delta} = j \mid X_t = i\right) \\
&- \Pr\left(X_{t+\Delta} = \tilde{j} \mid X_t = i\right) \Pr\left(X_{t+\Delta} = j \mid X_t = \tilde{i}\right) \quad \geq 0
\end{aligned}
\tag{17}
$$

for all quadruplets of states such that $i < \tilde{i}$ and $j < \tilde{j}$, and all real $\Delta > 0$ (by stationarity, the probabilities above are independent of t). The inequality is strict if we further assume that $\Pr(X_{t+\Delta} = j \mid X_t = i) > 0$ for every pair of states (i,j) and every $\Delta > 0$. Then we have:[8]

Proposition 5: Condition (17) is equivalent to the restriction that X can only jump from a given state to one of the two immediately adjacent states.

That is, for every state i, there exists $0 \leq \lambda_i \leq 1$ such that:[9]

[8] See Karlin and McGregor (1959a) for birth and death Markov chains.

$$\Pr\left(Y_{n+1} = j \mid Y_n = i\right) = \begin{cases} \lambda_i & \text{if } j = i+1 \\ 1 - \lambda_i & \text{if } j = i-1 \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

For instance, the example in Figure 5 violates the continuity condition (17) at jump times J_2 and J_3.

B. Discrete-Time, Discrete-State Trees

If we not only discretize the state space but also discretize the time dimension, then the natural representation of the dynamics of the process takes the form of a tree. Let Y_n in this case denote the state of the process after n moves, i.e., $Y_n = X_{n\Delta}$. In full generality, a tree is multinomial so that if $Y_n = i$ is the state after n moves, then Y_{n+1} can take any one of the possible states. As in the continuous-time Markov chain with transitions described by equation (18), continuity now means that the only possible moves occur to the immediately adjacent states or, in tree parlance, nodes. That is, binomial and trinomial trees are the natural approximation of a diffusion since by construction they restrict moves to take place to the immediately adjacent nodes. Conversely, binomial and trinomial trees with branches spaced with the same order of magnitude *cannot* approximate discontinuous processes such as jump-diffusions and more general Lévy processes despite the fact that they are commonly used in practice to price derivatives written on assets with discontinuous price dynamics. Basically, a jump is a move by one than one state at a time and, from what precedes, allowing for the possibility of jumps requires non-zero probabilities of moves to non-adjacent nodes (see Figure 6). I will explore below the implications of this for option pricing models based on binomial trees.

III. Some Examples

For now, I return to the continuous-state case of diffusions and give a few examples illustrating the applicability of the criterion function (15) to discriminate between diffusion and non-diffusion continuous-time models on the basis of their discrete-time transition functions.

---

[9] Note that λ_i is independent of n by stationarity.

A. *Example 1: Brownian Motion vs. Cauchy Jump Process*

The simplest possible illustration of the applicability of criterion (15) to distinguish a diffusion from a non-diffusion Markov process is provided by contrasting a Brownian motion, which has continuous sample paths, with a Cauchy process, which jumps.[10] The Brownian motion density

$$p(\Delta, y \mid x) = (2\pi\sigma^2\Delta)^{-1/2} \exp\left\{-(y-x)^2/(2\sigma^2\Delta)\right\} \tag{19}$$

satisfies (15), since $\partial^2 \mathrm{Ln}(p(\Delta, y \mid x))/\partial x \partial y = 1/(\sigma^2\Delta) > 0$. However, the Cauchy density

$$p(\Delta, y \mid x) = (\Delta/\pi)\big/\big((y-x)^2 + \Delta^2\big) \tag{20}$$

does not satisfy (15) --nor, consequently, (7)-- since

$$\frac{\partial^2 \mathrm{Ln}(p(\Delta, y \mid x))}{\partial x \partial y} = 2\left(\frac{\Delta^2 - (y-x)^2}{\Delta^2 + (y-x)^2}\right) < 0 \tag{21}$$

for y sufficiently far apart from x: $|y - x| > \Delta$. However, being Markov processes, they of course both satisfy the Chapman-Kolmogorov equation (2). They also satisfy the conservation requirement (6). Figure 7 plots the transition densities and the diffusion criterion (15) for these two models.

B. *Example 2: Variance Gamma Lévy Process*

This example illustrates that the approach applies to all non-diffusion processes – not just the usual Poisson jumps, but also more general Lévy jump processes. The Variance Gamma Lévy process is obtained by evaluating a Brownian motion at a random time with Gamma distribution. The result is a pure jump Lévy process. To create a model for stock prices, let $S_t = S_0 \exp(\mu t + VG_t)$ where $VG_t = BM_{\gamma_t}$. BM_t is an arithmetic Brownian motion with drift θ and diffusion σ^2, and γ_t is a random variable with a Gamma distribution with mean one and variance ν. The transition density of $X_t = \mathrm{Ln}(S_t)$ given X_0 is given by (see Madan, Carr and Chang (1998)):

---

[10] While both models have time-homogenous transition densities, neither model is stationary.

$$p\left(\Delta, y \mid x\right) = \frac{2 \exp\left(z(y\mid x)\theta/\sigma^2\right)}{\sqrt{2\pi}\sigma v^{\Delta/v}\Gamma\left(\Delta/v\right)} \left(\frac{z(y\mid x)^2}{\theta^2 + 2\sigma^2/v}\right)^{\frac{\Delta}{2v}-\frac{1}{4}} \text{BesselK}_{\frac{\Delta}{v}-\frac{1}{2}}\left(\frac{\sqrt{z(y\mid x)^2(\theta^2 + 2\sigma^2/v)}}{\sigma^2}\right) \quad (22)$$

where $z(y\mid x) = y - x - \mu\Delta$ and BesselK_α is the modified Bessel function of the second kind of order α. Of course, this density violates the diffusion criterion.

IV. The Implied Model from Option Prices: Is It a Diffusion?

A. The Transition Density Implicit in Option Prices

I now turn to the empirical question of determining whether the dynamic model for the underlying asset that is implied by observed option prices could be a diffusion. Suppose that we are interested in pricing at date zero a derivative security written on a traded underlying asset with price process $\{X_t \mid t \geq 0\}$, and with payoff function $\Psi(X_\Delta)$ at some future date Δ. Let us assume for simplicity that the riskless rate r and the dividend yield δ paid by the asset are constant. It is well-known that when markets are dynamically complete, the only price of the derivative security that is compatible with the absence of arbitrage opportunities is

$$P_0 \;=\; e^{-r\Delta}E\left[\Psi(X_\Delta)\Big| X_0 = x_0\right] \;=\; e^{-r\Delta}\int_0^{+\infty}\Psi(x)p(\Delta, x\mid x_0)dx \qquad (23)$$

where $p(\Delta,x\mid x_0)$ is the transition function (or risk-neutral density, or state-price density) induced by the dynamics of the underlying asset price.

Throughout this section, whenever I refer to the implied dynamics, or the implied model, I refer to their risk-neutral version. Note however that because the risk-neutral and actual probability measures assign zero probability to the same events (they have the same null sets), the underlying asset does not jump under one set of probabilities if and only if it does not jump under the other. So if we do not reject the null hypothesis that the underlying asset is a diffusion under the risk-neutral measure, then it must also be a diffusion under the actual probability measure. Conversely, if we reject it under the risk-neutral measure, then the underlying asset must be allowed to jump under the actual probability measure as well.

The Black-Scholes option pricing formula is the prime example of equation (23), when the underlying asset is a diffusion with $\sigma(x) = \sigma\,x$, σ constant. The corresponding transition density is the lognormal density

$$p_{BS}(\Delta, x \mid x_0) = \frac{1}{\sqrt{2\pi\Delta}\sigma x} \exp\left\{-\left(Ln(x/x_0) - (r - \delta - \sigma^2/2)\Delta\right)^2 \Big/ \left(2\sigma^2\Delta\right)\right\} \qquad (24)$$

and so the integral in equation (23) can be evaluated explicitly for specific payoff functions, such as the call option's $\Psi(X_\Delta) = \max(0, X_\Delta - K)$ for a fixed strike price K:

$$H_{BS}(\Delta, K, x_0, \sigma) = e^{-r\Delta}\left\{F_\Delta\Phi(d_1) - K\Phi(d_2)\right\} \qquad (25)$$

where $F_\Delta = x_0 \exp\left\{(r - \delta)\Delta\right\}$ is the forward price for delivery of the underlying asset at date Δ and $d_1 = \left(Ln(K/x_0) + (\sigma^2/2)\Delta\right)\Big/(\sigma\Delta^{1/2})$, $d_2 = d_1 - \sigma\Delta^{1/2}$.

At this point, the common practice when pricing and hedging equity options is to describe the market prices of call options for a given maturity Δ as given by the parametric equation (25) except that the volatility parameter for that maturity is a smooth function $\sigma_{IMP}(K/F_\Delta)$ of the option's moneyness $M = K/F_\Delta$:

$$H(\Delta, K, x_0) = H_{BS}\left(\Delta, K, x_0, \sigma_{IMP}(K/F_\Delta)\right). \qquad (26)$$

The function $\sigma_{IMP}(K/F_\Delta)$ is often known as the "implied volatility smile." A direct differentiation of the basic no-arbitrage pricing equation (23) with respect to the strike price, yields the corresponding risk neutral density.[11]

In the present setup, the only transition function compatible with the observed option prices H must be

$$\begin{aligned}
p(\Delta, K \mid x_0) &= e^{r\Delta}\frac{\partial^2}{\partial K^2}H(\Delta, K, x_0) \\
&= e^{r\Delta}\frac{\partial^2}{\partial K^2}H_{BS}\left(\Delta, K, x_0, \sigma_{IMP}(K/F_\Delta)\right) \qquad (27) \\
&= e^{r\Delta}\left\{\frac{\partial^2 H_{BS}}{\partial K^2} + \frac{2}{F_\Delta}\frac{d\sigma_{IMP}}{dM}\frac{\partial^2 H_{BS}}{\partial K\partial\sigma} + \frac{1}{F_\Delta^2}\left(\frac{d\sigma_{IMP}}{dM}\right)^2\frac{\partial^2 H_{BS}}{\partial\sigma^2} + \frac{1}{F_\Delta^2}\frac{d^2\sigma_{IMP}}{dM^2}\frac{\partial H_{BS}}{\partial\sigma}\right\}
\end{aligned}$$

---

[11] See Banz and Miller (1978) and Breeden and Litzenberger (1978), and Aït-Sahalia and Lo (1998) for a nonparametric version.

B. *Telling Whether the Implied Dynamic Model is a Diffusion*

I now turn to an empirical implementation of formula (27) with option price data. Option prices, or equivalently their implied volatilities, give us the function H in equation (26). Then equation (27) gives us the transition function for one maturity Δ implicit in the cross-section of option prices at one point in time. I will then check whether this transition function is compatible with an underlying diffusion model for the asset price by applying the criterion (15) to that implied transition function.

The data come from the Chicago Board Options Exchange (CBOE) and represent price quotes for call and put options on the Standard & Poor's 500 Index (SPX). The options are European, and to illustrate the methodology, I will focus on a single randomly chosen trading day, March 19, 2001. I repeated the experiment on different days, drawn from different time periods, to insure the robustness of the findings; the results are similar. I report the results for the most complete cross-section of traded strikes that day, the June 2001 expiration. Table I contains the full dataset used in the empirical application.

The raw data present three challenges. First, future dividends are not observable; second, S&P 500 futures are traded on the Chicago Mercantile Exchange, and cannot easily be time-stamped synchronously with the options to obtain F_Δ; and third, there are often substantial differences in the traded volume and open interest in the call and put with the same strike and maturity, except near the money where both are usually very liquid. I solve the first problem by relying on the spot-forward parity relationship under which the left hand side of equation (26) depends on the dividend yield δ only through F_Δ. To solve the second problem, I use prices of at-the-money options, where both the put and call are liquid, to infer the value of the implied futures F_Δ according to put-call-parity:

$$F_\Delta = K + e^{-r\Delta}\left\{H(\Delta,K,x_0) - G(\Delta,K,x_0)\right\} \tag{28}$$

where G denotes the put price and K is the strike closest to being at-the-money. Note that this equation does not require that the spot price of the index be recorded; it simply requires the market prices of the at-the-money call and put. Given the futures price F_Δ, I then replace the prices of all illiquid options, with the price implied by put-call parity applied at each value of the strike price, using the price of the more liquid option. For instance, if the put is more liquid then the call price is inferred from that of the put with the same strike as $H(\Delta,K,x_0) = G(\Delta,K,x_0) + e^{r\Delta}(F_\Delta - K)$ (this equation is instead solved for G given H when the call is more liquid). This solves the third problem. After this procedure, all the information

contained in liquid put prices has been extracted and resides in corresponding call prices. I can now concentrate exclusively on call options.

The first step consists in estimating the implied volatility function $\sigma_{IMP}(K/F_\Delta)$ in equation (26) . The data reveal quite clearly that an appropriate model for the implied volatility smile is a simple third order polynomial:

$$\sigma_{IMP}(K/F_\Delta) = \beta_0 + \beta_1(K/F_\Delta) + \beta_2(K/F_\Delta)^2 + \beta_3(K/F_\Delta)^3. \tag{29}$$

In fact, Table II shows that this model fits the data so well ($R^2 = 0.99$!) that one cannot help but wonder whether self-fulfilling prophecies are at play here. The same is true of other trading days and maturities in this market: the estimated parameters change, but not the quality of the fit. These options do indeed appear to "trade on a curve," the curve which is displayed in the top part of Figure 8. In the same way that in the 1970s and pre-1987 1980s the Black-Scholes model in its standard constant volatility form was a reasonably well-accepted paradigm, the current pricing model appears to be well-represented by equation (29) or slight variations of it, including for instance slightly different definitions of moneyness. The bottom plot in Figure 8 reports the estimated transition function from equation (27) corresponding to the implied volatility smile in the top plot.

I then apply the diffusion criterion (15) to test whether the underlying continuous-time model that produced the observed discrete-time transition represented by the function $p(\Delta, K \mid x_0)$ could have been a diffusion. Figure 9 displays the criterion function $\partial^2 Ln(p(\Delta, K \mid x_0))/\partial x_0 \partial K$, which gives the answer: the criterion function is negative in places, hence the underlying model for the S&P 500 index cannot be a diffusion.

Could the sampling noise, introduced by the fact that we need to estimate the parameters $\{\sigma_i \mid i = 0,...,3\}$ in equation (29), be sufficient to overturn the rejection of the diffusion hypothesis? The fact that model (29) describes the data in an accurate yet parsimonious way suggests that this is unlikely, but let us verify this formally. The effect of the estimation of the implied volatility parameters in equation (29) on the accuracy of the transition density estimator resulting from equation (27) can be assessed by the delta method. The parameter vector β is estimated using a sample of size n from the regression $\sigma = \sigma_{IMP}(K/F_\Delta) + \varepsilon$, of implied volatilities σ on moneyness K/F_Δ; ε is white noise with variance s^2 and M denote the vector of observed K/F_Δ. The distribution of the parameter estimates $\hat{\beta}$ is $n^{1/2}(\hat{\beta} - \beta) \rightarrow N(0, V_\beta)$, where $V_\beta = s^2(M'M)^{-1}$. From this it follows that

$$n^{1/2} \left(\frac{\partial^2 \text{Ln}(\hat{p}(\Delta, K \mid x_0))}{\partial K \partial x_0} - \frac{\partial^2 \text{Ln}(p(\Delta, K \mid x_0))}{\partial K \partial x_0} \right) \xrightarrow[n \to \infty]{d} N\left(0, D'.V_\beta.D\right)$$

where D denotes the gradient of $\partial^2 \text{Ln}(p(\Delta, K \mid x_0))/\partial x_0 \partial K$ with respect to the parameter vector β, and D' the transposed vector. D is easily calculated from equation (27).

A formal test can be based on calculating the minimum value reached by the criterion function $\partial^2 \text{Ln}(p(\Delta, K \mid x_0))/\partial x_0 \partial K$ over the interval of traded strikes $\left(\underline{K}, \overline{K}\right)$. Define this minimum, as a function of the parameters of the implied volatility smile, to be $\lambda(\beta)$. Let $\kappa(\beta)$ be the strike level at which this minimum is reached, i.e., the solution of the first order condition for the minimization of the criterion function. The minimum value is $\lambda(\beta) \equiv$ $\partial^2 \text{Ln}(p(\Delta, \kappa(\beta) \mid x_0))/\partial x_0 \partial K$. Keep differentiating each operation with respect to β to get to the gradient $\nabla \lambda$ of the minimum $\lambda(\beta)$ with respect to β. Then again by the delta method,

$$n^{1/2} \left(\lambda(\hat{\beta}) - \lambda(\beta) \right) \xrightarrow[n \to \infty]{d} N\left(0, \nabla \lambda'.V_\beta.\nabla \lambda\right). \tag{30}$$

Given estimates $\hat{\beta}$ and $\hat{V}_\beta$ of the implied volatility smile, we calculate $\hat{\lambda} \equiv \lambda(\hat{\beta})$ (negative otherwise there is no rejection) and $\hat{V}_\lambda \equiv \nabla \lambda'.\hat{V}_\beta.\nabla \lambda$ by evaluating the gradient at $\hat{\beta}$. Then the probability that the diffusion null hypothesis is true in the one-sided test of $H_0 : \lambda(\beta) \geq 0$ against $H_1 : \lambda(\beta) < 0$ is given by

$$\text{Prob}\left(H_0 \text{ true} \mid \hat{\lambda}, \hat{V}_\lambda\right) = \text{Prob}\left(\lambda > 0 \mid \hat{\lambda}, \hat{V}_\lambda\right) = 1 - \Phi\left(-n^{1/2}\hat{V}_\lambda^{-1/2}\hat{\lambda}\right). \tag{31}$$

where Φ is the standard Normal cumulative distribution function.

Not surprisingly, I find empirically that this noise is insufficient by itself to account for the negativity of the criterion function, i.e., the probability (31) is essentially zero, with the minimum value $\hat{\lambda}$ being negative and approximately ten times larger than its standard error $n^{-1/2}\hat{V}_\lambda^{1/2}$. The rejection of the diffusion hypothesis follows.[12] Given the quality of the fit of the implied volatilities, it is not surprising that the sampling noise is not sufficient to overturn the

---

[12] The situation would be quite different if the implied volatility model (29) were nonparametric, since we would then be subjected to the curse of differentiation in equation (27) and the sampling noise would be substantially greater. But given the fit of the basic parametric model, there seems to be no need for additional flexibility when modeling implied volatilities, at least with S&P 500 index options during that time period.

rejection, i.e., to make the negativity of $\partial^2 \mathrm{Ln}(p(\Delta, K \mid x_0))/\partial x_0 \partial K$ become statistically insignificant.

One last remark. Recall that the test for diffusion relies on a one factor assumption for the underlying asset returns. Could it be that rejecting that the underlying asset returns came from a diffusion is due to that univariate maintained hypothesis? The answer is no. Suppose that the model for the underlying asset price involves two factors, X as before and Y. For concreteness, think of Y as representing either another asset or X's stochastic volatility. Let $p(\Delta, x, y \mid x_0, y_0)$ be the corresponding transition density. The price of a derivative contract with payoff dependent on X_Δ is

$$
\begin{aligned}
P_0 &= e^{-r\Delta} E\left[\Psi(X_\Delta) \Big| X_0 = x_0, Y_0 = y_0 \right] \\
&= e^{-r\Delta} \int_{-\infty}^{+\infty} \int_0^{+\infty} \Psi(x) p(\Delta, x, y \mid x_0, y_0) dx dy \\
&= e^{-r\Delta} \int_{-\infty}^{+\infty} \int_0^{+\infty} \Psi(x) \left\{ \int_{-\infty}^{+\infty} p(\Delta, x, y \mid x_0, y_0) dy \right\} dx \\
&= e^{-r\Delta} \int_0^{+\infty} \Psi(x) p(\Delta, x \mid x_0, y_0) dx
\end{aligned}
$$

where now $p(\Delta, x \mid x_0, y_0) \equiv \int_{-\infty}^{+\infty} p(\Delta, x, y \mid x_0, y_0) dy$ represents the marginal in x from the conditional density of (x,y) given (x_0, y_0). Since the conditioning information is irrelevant in equation (27), the density extracted by the method of equation (27) should therefore be $p(\Delta, x \mid x_0, y_0)$ instead of $p(\Delta, x \mid x_0)$.

Hence if there were an additional factor, it should appear in the conditioning set of that density. For that, it should appear in the pricing formula H. Given that H is given by equation (26), the only place an additional factor could enter would be in the function σ_{IMP}, i.e., in the market-driven implied volatility smile. So to the extent that the market prices options using a deterministic smile model (the function σ_{IMP} which does not depend upon a second factor), and the evidence suggests that this is an accurate description of the reality (recall that $R^2 = 0.99$ for model (29)), the transition density implied by the market data will be a function of (x, x_0) only, not (x, x_0, y_0). By contrast, it is clear that trying to model the evolution in time of the transition density would undoubtedly require additional factors.

C. Consequences for Approximations of the Dynamics of the Underlying Asset and Implied Models

Dupire (1994) showed that, *if* the underlying model is a diffusion, then the call pricing function $H_t(\Delta, K, x_0)$ for maturity Δ at instant t necessarily satisfies the following forward form of the no-arbitrage pricing partial differential equation

$$\frac{\partial H_t}{\partial \Delta} = -\{r - \delta\} K \frac{\partial H_t}{\partial K} + \frac{1}{2} \frac{\partial^2 H_t}{\partial K^2} \sigma(K,t)^2 K^2 - \delta H_t \qquad (32)$$

from which it follows immediately that the "implied volatility function" compatible with the call pricing function is

$$\sigma^2(K,t) = \left\{\frac{\partial H_t}{\partial \Delta} + \{r - \delta\} K \frac{\partial H_t}{\partial K} + \delta H_t\right\} \bigg/ \left\{\frac{1}{2} \frac{\partial^2 H_t}{\partial K^2} K^2\right\}. \qquad (33)$$

Dumas, Fleming and Whaley (1998) used this methodology to empirically test different specifications of the function σ^2 against option data.

The analysis of the preceding section shows that, in light of the options data, there exists no such implied volatility function no matter how general it is allowed to be, since the underlying model is not a diffusion. In other words, the right hand side of (33) will not produce a function of (K,t) only. An appropriate modification of equation (32) allowing for discontinuities in the sample paths of the underlying asset would have to include difference terms (in addition to, or instead of, the differential terms).

Derman and Kani (1994) and Rubinstein (1994) propose binomial and trinomial tree approximations to the risk-neutral dynamics of the underlying asset that, by construction, replicate the observed option prices. Combining the empirical result of this section –that the underlying model is not a diffusion—with the characterization of tree approximations for continuous path processes in Section II.B, the conclusion is that these implied trees should be extended to more than three path possibilities at each node in order to be approximate discontinuous price paths. Otherwise, their continuous-time limit is constrained to be a continuous path process which we have seen is not the case empirically.

Therefore, contrary to the common practice in derivative pricing, jump-diffusion processes should be approximated by multinomial trees with more than three branches at least at some of their nodes. Of course, a binomial or trinomial tree could approximate a jump process

but only if at least one of the branches leads to a discrete change in the asset value, i.e., a change of order 1. This is not usually the case when trees are used in practice: the branches are often equally spaced, or at least the price changes from one node to the next over a time interval of length Δt are continuous in magnitude, i.e., of order $\sqrt{\Delta t}$.

D. Structural Alternatives to Implied Volatility Smiles

Modeling option prices by an implied volatility smile is inherently a purely descriptive approach. It is a very accurate description of actual market prices, but nevertheless remains a reduced-form approach. One of the main structural alternatives to the implied smile approach consists in extending the Black-Scholes in a variety of possible directions, which all involve relaxing the Normality assumption for the underlying asset returns that is built into equation (24) for the price density.

Consider first the ad hoc Edgeworth expansions that have been proposed in the literature.[13] These expansions replace the Normal density for returns with an Edgeworth expansion that allows for excess skewness and kurtosis in risk-neutral asset returns. If the stock price is S, the riskless rate r, σ the standard deviation of the stock return, and μ_3 and μ_4 denote the standardized skewness and kurtosis respectively, these expansions for the log-returns are typically in the form

$$p(\Delta, x \mid x_0) = \frac{\exp\{-z^2/2\}}{\sqrt{2\pi}} \left(1 + \frac{\mu_3}{6}\left(z^3 - 3z\right) + \frac{\mu_4}{24}\left(z^4 - 6z^2 + 3\right)\right) \tag{34}$$

where

$$z = z(\Delta, x \mid x_0) = \frac{x - x_0 - \left(r - \sigma^2/2\right)\Delta}{\sigma\sqrt{\Delta}}$$

with $x = Ln(S_\Delta)$ and $x_0 = Ln(S_0)$. Given equation (34), option pricing formulae can be obtained by applying equation (23) with the density p for prices replaced by its implication from the returns density (34).

Note that the transition density $p(\Delta, x \mid x_0)$ in this model is a function of (x, x_0) only through $z = x - x_0$. Note that, beyond option pricing, this form of space-homogeneity has

---

[13] See for example Jarrow and Rudd (1982).

important statistical consequences when combined with time-homogeneity. Even if the model is not stationary, its first differences are, thereby making the analysis of maximum-likelihood estimators substantially simpler. Indeed, treating the first observation as fixed, the likelihood function is

$$\ell_n\left(\theta\right) \equiv \sum_{i=1}^{n} \text{Ln}\left\{p_\theta\left(\Delta, r_{i\Delta} \mid r_{(i-1)\Delta}\right)\right\} = \sum_{i=1}^{n} \text{Ln}\left\{p_\theta\left(\Delta, r_{i\Delta} - r_{(i-1)\Delta}\right)\right\} \tag{35}$$

and in the latter form involves only the data $(r_{i\Delta} - r_{(i-1)\Delta})$.

But could equation (34) represent the transition density of a diffusion? In other words, we need to find out what are the diffusions with space-homogeneous transition functions. Using the diffusion criterion, it turns out that the class of such processes is rather small:[14]

Proposition 6: The only diffusion with a space-homogenous transition function is the (arithmetic) Brownian motion, i.e., the process with $\sigma(x) = \sigma$ and $\mu(x) = \mu$ both constant.

The Edgeworth expansion (34) is in the space-homogenous class $p(\Delta,x|x_0) = q(\Delta,x-x_0)$. But from Proposition 6 the Gaussian distribution is the only diffusion process with a space-homogeneous transition function $q(\Delta,x-x_0)$. Hence there is no diffusion model for the underlying stock returns that can be represented by these ad hoc expansions, other than the Gaussian density for which $\mu_3 = \mu_4 = 0$. Note that this is not just saying that equation (34) is not, in general, a proper density. There is just no density, other than the Gaussian, that can represent a diffusion and be space-homogeneous.[15]

This leads quite naturally to the next idea. How about maintaining the Gaussian assumption but making the conditional mean and variance more complex than in the Black-Scholes model? Could that accommodate a diffusion process? Unfortunately, the answer is, here too, negative. Suppose that we restrict attention to processes with Gaussian distributions, and use

---

[14] The result can be obtained differently by calculating the Laplace transform of a Polya frequency density of order 2 (see Karlin (1968, Chapter 7) for definitions and Theorem 5.2).

[15] Since the transition density of the Variance Gamma Lévy process is space-homogeneous but not Gaussian, it cannot represent the transition function of a diffusion. However, the function $p(\Delta,K|x_0)$ produced by equation (27) is not a function of $K - x_0$. Hence Proposition 6 cannot be applied to rule out immediately the possibility of an underlying diffusion model when using an implied volatility smile. We must use the diffusion criterion (15) to be able to tell.

the results here to prove differently a well-known result regarding Markov processes with Gaussian transitions. Namely, we have:

Proposition 7: The Ornstein-Uhlenbeck process

$$dX_t = \left(\alpha - \beta X_t\right)dt + \sigma dZ_t \tag{36}$$

is not only the only diffusion with a Gaussian transition function, but also the only such Markov process.

The conditional mean and variance of the process are affine and constant respectively, $E[X_{t+\Delta} \mid X_t = x] = \alpha/\beta + (x - \alpha/\beta)e^{-\beta\Delta}$ and $V[X_{t+\Delta} \mid X_t = x] = \sigma^2(1 - e^{-2\beta\Delta})/2\beta$, which reduce to $E[X_{t+\Delta} \mid X_t = x] = x$ and $V[X_{t+\Delta} \mid X_t = x] = \sigma^2\Delta$ if $\beta=0$ (the arithmetic Brownian motion special case). The implication of this for potential structural extensions of the Black-Scholes model wishing to remain within the diffusion class is that, in the same fashion that the model cannot be extended to non-Gaussian but space-homogenous densities, it cannot be extended either to Gaussian densities with conditional mean and variances that are more complex than affine and constant respectively. The latter extension would in fact take us even outside of the Markov class in which case pricing via risk-neutral expectations conditioned on the current asset price given by equation (23) no longer holds.

V. Conclusions

Within the Markov world, diffusion processes are characterized by the continuity of their sample paths. When looking at discrete data, are the discontinuities observed the result of discreteness, or are they the result of non-diffusion behavior on the part of the underlying continuous-time data-generating process? This paper examined the implications for the discrete data of having been generated by a univariate diffusion, on the basis of a criterion that uniquely characterizes the transition densities of diffusions, and is equivalent to continuity of the continuous-time, unobservable, sample paths. It relies solely on the transition function, an object that can be inferred from the discrete observations.

I also interpreted this characterization in terms of discrete-state processes, first continuous-time Markov chains and then discrete-time trees. The intuitive result is that the characterization is this case means that the process can only jump by one state at a time. I then drew some implications for the approximations used in derivative pricing, depending upon whether the underlying model is or is not a diffusion. Finally, I tested whether the underlying

model for the asset price dynamics that is implied by S&P 500 option prices could have been a diffusion, and spelled out some implications for the implied diffusion, implied tree and Edgeworth expansions approaches to option pricing.

One final remark. From a discretely sampled time-series $\{r_0, r_\Delta, r_{2\Delta}, \ldots, r_{n\Delta}\}$, one could test nonparametrically the hypothesis that the data were generated by a continuous-time diffusion $\{r_t, t \geq 0\}$. Formally,

$$
\begin{cases}
H_0 : \partial^2 Ln(p(\Delta, y \mid x)) / \partial x \partial y > 0 & \text{for all } x, y \\
H_1 : \partial^2 Ln(p(\Delta, y \mid x)) / \partial x \partial y \leq 0 & \text{for some } x, y
\end{cases}
$$

and one could base a test on checking whether (15) holds for a nonparametric estimator of the density $p(\Delta, y \mid x)$. Locally polynomial estimators can be used for that purpose (see Aït-Sahalia (2000)). Their use in the present context is left to future work.

References

Ahn, Chang Mo and Howard E. Thompson, 1988, Jump-diffusion processes and the term structure of interest rates, *Journal of Finance* 43, 155-174.

Aït-Sahalia, Yacine, 1996a, Nonparametric pricing of interest rate derivative securities, *Econometrica* 64, 527-560.

Aït-Sahalia, Yacine, 1996b, Testing continuous-time models of the spot interest rate, *The Review of Financial Studies* 9, 385-426.

Aït-Sahalia, Yacine, 1997, Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach, *Econometrica,* forthcoming .

Aït-Sahalia, Yacine, 1999, Transition densities for interest rate and other nonlinear diffusions, *Journal of Finance* 54, 1361-1395.

Aït-Sahalia, Yacine, 2000, Testing the Markov hypothesis in discretely-sampled continuous-time processes, Working paper, Princeton University.

Aït-Sahalia, Yacine, and Andrew Lo 1998, Nonparametric estimation of state-price-densities implicit in financial asset prices, *Journal of Finance* 53, 499-547.

Aït-Sahalia, Yacine, Yubo Wang, and Francis Yared, 2001, Do option markets correctly price the probabilities of movement of the underlying asset?, *Journal of Econometrics* 102, 67-110. .

Banz, Rolf, and Merton Miller, 1978, Prices for state- contingent claims: Some estimates and applications, *Journal of Business* 51, 653-672.

Breeden, Douglas, and Robert H. Litzenberger, 1978, Prices of state-contingent claims implicit in option prices, *Journal of Business* 51, 621-651.

Bates, David, 1991, The crash of '87: Was it expected? The evidence from options markets, *Journal of Finance* 46, 1009-1044.

Cox, John C., John E. Ingersoll, and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385-407.

Das, Sanjiv R. and Silverio Foresi, 1996, Exact solutions for bond and option prices with systematic jump risk, *Review of Derivatives Research* 1, 7-24.

Duffie, Darrell, Jun Pan, and Kenneth Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* 68, 1343-1376.

Dumas, Bernard, Jeffrey Fleming, and Robert E. Whaley, 1999, Implied volatility functions: Empirical tests, *Journal of Finance* 53, 2059-2106.

Derman, Emanuel, and Iraj Kani, 1994, Riding on the smile, *RISK* 7, 32-39.

Dupire, Bruno, 1994, Pricing with a smile, *RISK* 7, 18-20.

Feller, William, 1954, The general diffusion operator and positivity preserving semi-groups in one dimension, *Annals of Mathematics* 60, 417-436.

Feller, William, 1971, *An Introduction to Probability Theory and Its Applications, Volume II,* Second Edition (John Wiley & Sons, New York).

Florens, Jean-Pierre, Eric Renault, and Nizar Touzi, 1998, Testing for embeddability by stationary scalar diffusions, *Econometric Theory* 14, 744-769.

Friedman, Avner, 1975, *Stochastic Differential Equations and Applications, Volume 1* (Academic Press, New York).

Gikhman, I.I. and A.V. Skorohod, 1969, *Introduction to the Theory of Random Processes* (W.B. Sauders Company, Philadelphia).

Hansen, Lars Peter, José A. Scheinkman and Nizar Touzi, 1998, Identification of scalar diffusions using eigenvectors, *Journal of Econometrics*, 86, 1-32.

Jarrow, Robert and Andrew Rudd, 1982, Approximate option valuation for arbitrary stochastic processes, *Journal of Financial Economics* 10, 347-369.

Karlin, Samuel, 1968, *Total Positivity*, Volume I (Stanford University Press, Stanford).

Karlin, Samuel and James McGregor, 1959a, Coincidence properties of birth-and-death processes, *Pacific Journal of Mathematics* 9, 1109-1140.

Karlin, Samuel and James McGregor, 1959b, Coincidence probabilities, *Pacific Journal of Mathematics* 9, 1141-1164.

Kou, Steven, 1999, A jump diffusion model for option pricing with three properties: leptokurtic feature, volatility smile, and analytical tractability, Working paper, Columbia University.

Madan, Dilip B., Peter P. Carr and Eric C. Chang, 1998, The variance Gamma process and option pricing, *European Finance Review* 2, 79-105.

Mandl, Petr, 1968, *Analytical Treatment of One-Dimensional Markov Processes* (Springer-Verlag, Berlin).

Merton, Robert C., 1976, Option pricing when underlying stock returns are discontinuous, *Journal of Financial Economics* 3, 125-144.

Norris, James R., 1997, *Markov Chains* (Cambridge University Press, Cambridge, U.K.).

Ray, Daniel, 1956, Stationary Markov processes with continuous paths, *Transactions of the American Mathematical Society* 82, 452-493.

Rubinstein, Mark, 1994, Implied binomial trees, *Journal of Finance* 49, 771-818.

Vasicek, Oldrich, 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177-188.

Appendix

Proof of Proposition 1

To prove this claim, it suffices to show that (14) implies that the convergence in (7) is uniform in x , which is known to be a necessary and sufficient condition for the continuity of the sample paths. Let I be a compact interval included the domain D=($\underline{r}$, $\overline{r}$) of the process r. That (7) holds for every fixed $\varepsilon > 0$ and fixed x in I is fairly innocuous. Indeed, assume (this is a mild additional requirement) that the convergence in the limit (6) occurs at an arbitrary polynomial rate, that is

$$P\left(\Delta,(\underline{r},x-\varepsilon)\cup(x+\varepsilon,\overline{r})\mid x\right)=1-\int_{|y-x|\leq\varepsilon}p\left(\Delta,y\mid x\right)dy \ \leq\ c_{\varepsilon,x}\Delta^{\kappa_{\varepsilon,x}} \tag{37}$$

for some constants $c_{\varepsilon,x}>0$ and $\kappa>0$ (κ is not necessarily greater or equal to 1). Then it follows from (13) that

$$P\left(\Delta,(x+\varepsilon,\overline{r})\mid x\right)P\left(\Delta,(x+\varepsilon/4,x+3\varepsilon/4)\mid x+\varepsilon/2\right)$$
$$\leq P\left(\Delta,(x+\varepsilon/4,x+3\varepsilon/4)\mid x\right)P\left(\Delta,(x+\varepsilon,\overline{r})\mid x+\varepsilon/2\right) \tag{38}$$

From (37), we have that

$$\begin{cases} P\left(\Delta,(x+\varepsilon/4,x+3\varepsilon/4)\mid x\right) \ \leq\ c_{\varepsilon/4,x}\Delta^{\kappa} \\ P\left(\Delta,(x+\varepsilon,\overline{r})\mid x+\varepsilon/2\right) \ \leq\ c_{\varepsilon/2,x+\varepsilon/2}\Delta^{\kappa} \end{cases}$$

and, since from (6) there exists $\zeta>0$ such that $P\left(\Delta,(x+\varepsilon/4,x+3\varepsilon/4)\mid x+\varepsilon/2\right) \geq 1/2$ for all $0\leq\Delta\leq\zeta$, it follows from (38) that $P\left(\Delta,(x+\varepsilon,\overline{r})\mid x\right) \leq\ 2\,d_{x,\varepsilon}\,\Delta^{2\kappa}$ for some constant $d_{\varepsilon,x}>0$. This process can be repeated an arbitrary number of times, each time multiplying the constant κ by a factor of 2. Therefore the constant κ can be replaced by 1, or for that matter any number, but 1 will be sufficient in what follows.

The crucial aspect now is to prove the uniformity of the convergence in x. Fix $\varepsilon>0$ and let $x_0 < x_1 < \ldots < x_m$ be a finite partition of I, such that $x_i - x_{i-1} = \varepsilon/2$ for i=1,...,m. Fix $\tilde{\varepsilon}>0$. From what precedes wit κ now replaced by 1, for each i=1,...,m, there exists $\xi_i>0$ such that for all Δ, $0\leq\Delta\leq\xi_i$ implies $P\left(\Delta,(x_i+\varepsilon/2,\overline{r})\mid x_i\right) \leq(\tilde{\varepsilon}/2)\Delta$. For x in I, consider the particular x_i such that $x-\varepsilon/2 < x_{i-1} \leq x \ < x_i < x+\varepsilon/2$. For $0\leq\Delta\leq\xi_i$, it follows from (13) that

$$P\big(\Delta,(x+\varepsilon,\overline{r})\,|\,x\big)P\big(\Delta,(\underline{r},x+\varepsilon)\,|\,x_i\big) \le P\big(\Delta,(x+\varepsilon,\overline{r})\,|\,x_i\big)P\big(\Delta,(\underline{r},x+\varepsilon)\,|\,x\big)$$

$$\le P\big(\Delta,(x+\varepsilon,\overline{r})\,|\,x_i\big)$$

$$\le P\big(\Delta,(x_i+\varepsilon/2,\overline{r})\,|\,x_i\big) \tag{39}$$

$$\le (\tilde{\varepsilon}/2)\Delta$$

Now, from (6), there exists $\zeta_i>0$ such that $P\big(\Delta,(\underline{r},x_{i-1}+\varepsilon)\,|\,x_i\big)\ge 1/2$ for all $0\le\Delta\le\zeta_i$. Next, for Δ satisfying $0\le\Delta\le\inf\{\eta_i,\zeta_i\}$, we have that $P\big(\Delta,(x+\varepsilon,\overline{r})\,|\,x\big)\le\tilde{\varepsilon}\Delta$ since $P\big(\Delta,(\underline{r},x+\varepsilon)\,|\,x_i\big)\ge P\big(\Delta,(\underline{r},x_{i-1}+\varepsilon)\,|\,x_i\big)$. Now define $\eta_I\equiv\inf\big\{\{\xi_i,\zeta_i\}\,|\,i=1,\ldots,m\big\}$, which is independent of the particular x in I. Thus there exists $\eta_I>0$ such that for any $0\le\Delta\le\eta_I$, $P\big(\Delta,(x+\varepsilon,\overline{r})\,|\,x\big)\le\tilde{\varepsilon}\Delta$ for all x in I. Similarly, there exists $\chi_I>0$ such that for any $0\le\Delta\le\chi_I$, $P\big(\Delta,(\underline{r},x-\varepsilon)\,|\,x\big)\le\tilde{\varepsilon}\Delta$ for all x in I. Since

$$\int_{|y-x|>\varepsilon}p\big(\Delta,y\,|\,x\big)dy\ =P\big(\Delta,(\underline{r},x-\varepsilon)\,|\,x\big)+P\big(\Delta,(x+\varepsilon,\overline{r})\,|\,x\big)\ \le\ 2\tilde{\varepsilon}\Delta,$$

it follows that the convergence in (7) holds uniformly over x in the compact interval I.

Proof of Proposition 2

First recall that diffusions are Markov processes, hence they must satisfy the Chapman-Kolmogorov equation (2). Therefore

$$\delta\big(2\Delta,y,\tilde{y}\,|\,x,\tilde{x}\big)\ =\int_{\underline{r}}^{\overline{r}}\int_{\underline{r}}^{\overline{r}}p\big(\Delta,y\,|\,z\big)p\big(\Delta,z\,|\,x\big)p\big(\Delta,\tilde{y}\,|\,\tilde{z}\big)p\big(\Delta,\tilde{z}\,|\,\tilde{x}\big)dzd\tilde{z}$$

$$-\int_{\underline{r}}^{\overline{r}}\int_{\underline{r}}^{\overline{r}}p\big(\Delta,y\,|\,\tilde{z}\big)p\big(\Delta,\tilde{z}\,|\,\tilde{x}\big)p\big(\Delta,\tilde{y}\,|\,z\big)p\big(\Delta,z\,|\,x\big)dzd\tilde{z}$$

$$=\int_{\underline{r}}^{\overline{r}}\int_{\underline{r}}^{\overline{r}}\delta\big(\Delta,y,\tilde{y}\,|\,z,\tilde{z}\big)p\big(\Delta,z\,|\,x\big)p\big(\Delta,\tilde{z}\,|\,\tilde{x}\big)dzd\tilde{z}$$

$$=\int_{\underline{r}}^{\overline{r}}\int_{\underline{r},\ z<\tilde{z}}^{\overline{r}}\delta\big(\Delta,y,\tilde{y}\,|\,z,\tilde{z}\big)\big[p\big(\Delta,z\,|\,x\big)p\big(\Delta,\tilde{z}\,|\,\tilde{x}\big)$$

$$-p\big(\Delta,\tilde{z}\,|\,x\big)p\big(\Delta,z\,|\,\tilde{x}\big)\big]dzd\tilde{z}$$

where we have exploited the symmetry property that $\delta\big(\Delta,y,\tilde{y}\,|\,\tilde{z},z\big)=\ -\delta\big(\Delta,y,\tilde{y}\,|\,z,\tilde{z}\big)$ for any $(z,\tilde{z})$, so we only need to integrate over the half-quadrant where $z<\tilde{z}$. As a result

$$\delta\big(2\Delta,y,\tilde{y}\,|\,x,\tilde{x}\big)=\ \int_{\underline{r}}^{\overline{r}}\int_{\underline{r},\ z<\tilde{z}}^{\overline{r}}\delta\big(\Delta,y,\tilde{y}\,|\,z,\tilde{z}\big)\delta\big(\Delta,z,\tilde{z}\,|\,x,\tilde{x}\big)dzd\tilde{z}\ >0 \tag{40}$$

since (14) holds for all transitions of length Δ. The same argument shows that the property is satisfied for all integer multiples of Δ, i.e., for all observable frequencies.

Proof of Proposition 3

Consider $x < \tilde{x}$ and $y < \tilde{y}$. Rearranging inequality (14) yields the equivalent formulation

$$\frac{1}{\tilde{y} - y}\left(\left\{\frac{Ln(p(\Delta, \tilde{y} \mid \tilde{x})) - Ln(p(\Delta, \tilde{y} \mid x))}{\tilde{x} - x}\right\} - \left\{\frac{Ln(p(\Delta, y \mid \tilde{x})) - Ln(p(\Delta, y \mid x))}{\tilde{x} - x}\right\}\right) > 0. \quad (41)$$

The necessity of (15) then follows by taking the limit as $\tilde{x} \to x^+$ and $\tilde{y} \to y^+$ in (41). The sufficiency of (15) follows from the fact that a function whose derivative is positive is increasing: at a fixed x, $\partial Ln(p)/\partial x$ is an increasing function of y, hence $\partial Ln(p(\Delta, \tilde{y} \mid x))/\partial x > \partial Ln(p(\Delta, y \mid x))/\partial x$. Thus $\left\{Ln(p(\Delta, \tilde{y} \mid x)) - Ln(p(\Delta, y \mid x))\right\}/\left\{\tilde{y} - y\right\}$ is an increasing function of x. Then (41), or equivalently (14), follows.

Proof of Proposition 4

We want to show that, if condition (15) is satisfied by the transition density p_X of a process X, then the same condition is fulfilled by the transition density p_Y of the process $Y = \varphi^{-1}(X)$ where the function φ is twice continuously differentiable and strictly monotonic.[16] Indeed the transition density of Y is given by the Jacobian formula:

$$\begin{aligned} p_Y\left(\Delta, y \mid y_0\right) &= \frac{\partial}{\partial y} \text{Prob}\left(Y_{t+\Delta} \le y \mid Y_t = y_0\right) = \frac{\partial}{\partial y} \text{Prob}\left(X_{t+\Delta} \le \varphi(y) \mid X_t = \varphi(y_0)\right) \\ &= \frac{\partial}{\partial y}\left[\int^{\varphi(y)} p_X\left(\Delta, x \mid \varphi(y_0)\right) dx\right] = \varphi'(y)\, p_X\left(\Delta, \varphi(y) \mid \varphi(y_0)\right) \end{aligned} \quad (42)$$

Consequently, if X is a diffusion then $\delta_X\left(\Delta, x \mid x_0\right) \equiv \partial^2 Ln(p(\Delta, y \mid x))/\partial x \partial y > 0$ from condition (15) and it follows from equation (42) that

$$\delta_Y\left(\Delta, y \mid y_0\right) \equiv \frac{\partial^2}{\partial y_0 \partial y} Ln(p_Y(\Delta, y \mid y_0)) = \varphi'(y)\varphi'(y_0)\delta_X\left(\Delta, \varphi(y) \mid \varphi(y_0)\right) > 0 \quad (43)$$

so the transition function of the Y process automatically satisfies condition (15).

---

[16] Such transformations are commonly used in finance, for instance to go from an arithmetic Brownian motion to a geometric Brownian motion, from a Cox-Ingersoll-Ross square-root process to a Bessel process, from a CEV process (with geometric mean) to a Bessel process, etc. Indeed, most closed-form solutions we rely on are obtained through such a transformation.

Proof of Proposition 5

With $J_0, J_1, \ldots$ denoting the jump times of the chain $X = \{X_t\}_{t \geq 0}$, and $S_1, S_2, \ldots$ its holding times, defined by

$$S_n = \begin{cases} J_n - J_{n-1} & \text{if } J_{n-1} < \infty \\ \infty & \text{otherwise} \end{cases} \tag{44}$$

right-continuity forces $S_n > 0$ for all n. If $J_{n+1} = \infty$ for some n, define $X_\infty \equiv X_{J_n}$, otherwise X_∞ is undefined. For convenience, we set $X_t = \infty$ if t is greater than the first explosion time $\sup\{J_n / n \geq 0\} = \sum_{n=1}^{\infty} S_n$.

From the theory of continuous-time Markov chains (see e.g., Norris (1997), page 87), condition (18) on the transition matrix of the jump chain Y is equivalent to the restriction that the generator matrix A of the Markov chain X be of the Jacobi form: zero entries except on the diagonal, supra-diagonal and infra-diagonal lines, i.e.:

$$A = \begin{pmatrix} \ddots & \ddots & & \ddots & & \ddots \ddots \\ \cdots 0 & \beta_i & -\alpha_i - \beta_i & \alpha_i & 0 \cdots \\ & \ddots & \ddots & & \ddots & \ddots \ddots \end{pmatrix} \tag{45}$$

where $\alpha_i > 0$, $\beta_i > 0$ and then $\lambda_i = \alpha_i / (\alpha_i + \beta_i)$. The element $[a_{ij}]$ of A determines the rate at which the chain moves from state i to state j.

To prove that (17) is equivalent to (45), first note from (2) --where the integral is replaced by a sum over all the possible intermediary states-- that it suffices to prove the equivalence for an infinitesimal Δ, and the equivalence will then be carried forward in time by repeated use of (2). Next express the transition matrix $P(\Delta)$ in terms of the generator matrix A:

$$P(\Delta) = \exp[\Delta A] = \sum_{k=1}^{\infty} \frac{\Delta^k A^k}{k!}. \tag{46}$$

Suppose that A has a non-zero element outside the three lines indicated in (45), say $a_{i-1,i+1} = \gamma_{i-1} > 0$, with $a_{i-1,i-1} = -\alpha_{i-1} - \beta_{i-1} - \gamma_{i-1}$ now being the required diagonal term on the row. Then it follows from $P(\Delta) = I + \Delta A + o(\Delta)$ that (17) is violated since

$$\delta(\Delta, i, i+1 \mid i-1, i) = -\gamma_{i-1}\Delta + O(\Delta^2) \quad < 0 \tag{47}$$

By contrast, if A has the form (45), then it follows from

$$P(\Delta) = I + \Delta A + \Delta^2 A^2 / 2 + o(\Delta^2) . \tag{48}$$

that

$$\delta\left(\Delta, i, i+1 \mid i-1, i\right) = \alpha_i \alpha_{i-1} \Delta^2 - \alpha_i \alpha_{i-1} \Delta^2 / 2 + o(\Delta^2) = \alpha_i \alpha_{i-1} \Delta^2 / 2 \geq 0 \tag{49}$$

and similarly for the other transitions. The equivalence is therefore proved. Note that the form of the generator (45) characterizes birth and death processes, where the size of the population either goes up or down by one individual at a time. As a special case, this class includes Poisson processes for which $Y_n = n$ with probability one, so $\lambda_i = 1$ for all i.

Proof of Proposition 6

The criterion (15) in the case of a transition density depending only on z reduces to $\partial^2 \mathrm{Ln}(p(\Delta, z))/\partial z^2 < 0$, i.e., log-concavity of the density. A large number of densities are log-concave. But one must remember that the process must be Markovian, i.e., satisfy (2). The combination of both is enough to reduce the set of admissible models to the arithmetic Brownian motion.

Indeed, consider the leading term at order Δ^{-1} of the expansion of the transition density of such a space-homogeneous diffusion:

$$\mathrm{Ln}\left(\tilde{p}_X^{(-1)}\left(\Delta, x_0 + z \mid x_0\right)\right) = -\frac{1}{2}\mathrm{Ln}\left(2\pi\Delta\right) - \mathrm{Ln}\left(\sigma(x_0 + z)\right) - \frac{1}{2\Delta}\left(\gamma(x_0 + z) - \gamma(x_0)\right)^2 \tag{50}$$

(see Aït-Sahalia (1997, 1999)). We are asking when this function depends on z but not on x_0. By a Taylor expansion in z around 0, the right hand side of equation (50) is independent of x_0 if and only if the function σ is, in which case $\gamma(x) = x/\sigma$. Looking then at the next order term,

$$\mathrm{Ln}\left(\tilde{p}_X^{(0)}\left(\Delta, x_0 + z \mid x_0\right)\right) - \mathrm{Ln}\left(\tilde{p}_X^{(-1)}\left(\Delta, x_0 + z \mid x_0\right)\right) = \int_{\gamma(x_0)}^{\gamma(x_0 + z)} \mu_Y(v) dv \tag{51}$$

where

$$\mu_Y(y) = \mu(\gamma^{-1}(y))/\sigma(\gamma^{-1}(y)) - \sigma'(\gamma^{-1}(y))/2 . \tag{52}$$

Given that γ is linear, the right-hand-side of equation (51) can only be independent of x_0 if $\mu_Y(.)$ is constant. From equation (52) and $\sigma(x) = \sigma$, $\gamma(x) = x/\sigma$, this can only occur if $\mu(x) = \mu$ is constant. Hence the only process with space-homogeneous transition density is the arithmetic Brownian motion with $\sigma(x) = \sigma$ and $\mu(x) = \mu$ both constant.

Proof of Proposition 7

First note that if $(X_{t+\Delta}, X_t)$ has a Gaussian distribution, then so does $X_{t+\Delta} \mid X_t$ and moreover $E[X_{t+\Delta} \mid X_t]$ must be affine in X_t and $V[X_{t+\Delta} \mid X_t]$ constant in X_t. This follows for instance from Theorem III.6.5 page 86 of Feller (1971). Let us therefore consider Gaussian transition functions $p(\Delta, y \mid x)$ with conditional mean $e_\Delta(x) \equiv E[X_{t+\Delta} \mid X_t = x] = \alpha_\Delta + \beta_\Delta x$ and conditional variance $v_\Delta \equiv V[X_{t+\Delta} \mid X_t = x]$, and let us see what further restrictions on the dependence of α_Δ, β_Δ and v_Δ on Δ make these functions compatible with the conditions (2) for Markovianity and (15) for diffusion. That is, should further restrictions on the conditional mean and variance be placed to insure that the discrete observations are embeddable in a diffusion?

It turns out that in this case the Markov requirement alone is sufficient to reduce the set of compatible transition densities to the Ornstein-Uhlenbeck case which we already know is a diffusion.[17] In other words, there exists no non-diffusion Markov process with Gaussian transitions. Equivalently, this means that for Gaussian transitions once the Markov requirement is imposed, the diffusion condition (15) is automatically satisfied. Indeed, the Chapman-Kolmogorov equation (2) implies by a direct calculation that

$$\alpha_{2\Delta} = \alpha_\Delta (1 + \beta_\Delta), \ \beta_{2\Delta} = \beta_\Delta^2 \ , \ v_{2\Delta} = v_\Delta (1 + \beta_\Delta^2).$$

With $\alpha_0 = 0$, $\beta_0 = 1$ and $v_0 = 0$, the only solution is of the form

$$\alpha_\Delta = \alpha (1 - e^{-\beta\Delta}), \ \beta_\Delta = e^{-\beta\Delta} \ , \ v_\Delta = \frac{v}{2\beta} (1 - e^{-2\beta\Delta})$$

where α, β and $v > 0$ are constants. Applying now condition (15) to

$$p(\Delta, y \mid x) = (2\pi v_\Delta)^{-1/2} \exp\left\{-(y - e_\Delta(x))^2 / (2v_\Delta)\right\}$$

yields $\partial^2 \mathrm{Ln}(p(\Delta, y \mid x)) / \partial x \partial y = \beta_\Delta / v_\Delta > 0$. Therefore the diffusion condition puts no additional constraints on the model's parameters in the Gaussian case so that saying that the process is Markovian and Gaussian is enough to reduce the admissible set to the Ornstein-Uhlenbeck class.

---

[17] Note that this includes the arithmetic Brownian motion as the special case where $\beta = 0$.

Table I

SPX Option Data

Strike	Call Price	Moneyness	Implied Volatility
750	408.50	0.6449	0.4131
800	359.90	0.6879	0.3903
850	311.60	0.7309	0.3638
900	264.40	0.7739	0.3463
950	218.40	0.8169	0.3274
995	178.30	0.8556	0.3081
1025	153.00	0.8814	0.2978
1050	132.70	0.9029	0.2885
1075	113.60	0.9244	0.2805
1100	95.30	0.9459	0.2706
1125	78.60	0.9674	0.2624
1130	75.30	0.9717	0.2601
1140	69.30	0.9803	0.2575
1150	63.20	0.9889	0.2533
1160	57.60	0.9975	0.2502
1170	52.30	1.0061	0.2473
1175	49.60	1.0104	0.2451
1180	47.50	1.0147	0.2453
1190	42.50	1.0233	0.2412
1200	37.80	1.0319	0.2372
1210	33.60	1.0405	0.2341
1225	28.00	1.0534	0.2303
1250	20.35	1.0749	0.2253
1275	14.05	1.0964	0.2190
1300	9.50	1.1179	0.2144
1325	6.10	1.1394	0.2092
1350	3.85	1.1609	0.2057
1375	2.40	1.1824	0.2033
1400	1.475	1.2039	0.2017
1425	0.875	1.2254	0.1999
1450	0.475	1.2469	0.1968
1475	0.350	1.2684	0.2019
1500	0.225	1.2898	0.2035

These options are European calls on the S&P 500 index with prices recorded on March 19, 2001, at 10:30 AM CST. For each option's price, I use the bid-ask midpoint. The riskless rate is r=5.50% (which, following market convention, is slightly higher than the 3-month T-Bill rate, reflecting the fact that T-Bill rate is not the relevant riskless rate faced by traders). The options expire on June 15, 2001. With the calendar convention, these options have $\Delta = 88$ days to expiration. The at-the-money implied forward price of the index for that maturity is F_Δ=1,162.93 while the value of the index itself is X_0=1,151.10. Time calculations are performed with a 365-day calendar. The moneyness of an option with strike K is $M = K/F_\Delta$.

Table II

Fitted Implied Volatility Smile

| Variable	Coefficient	t-statistic	p-value
1	0.4775	10.02	6.10^{-11}
M	0.5221	3.42	2.10^{-3}
M^2	-1.3714	-8.59	2.10^{-9}
M^3	0.6208	11.39	3.10^{-12}

This table reports the results of fitting the model (29) to the implied volatility data given in Table I. M denotes the option's moneyness $M = K/F_\Delta$. The R^2 of the regression is 0.9993; the adjusted R^2 is 0.9992. The fitted implied volatility function $\sigma_{IMP}(K/F_\Delta)$ is plotted in Figure 8.

Figure 1: Discrete and Continuous Sample Paths
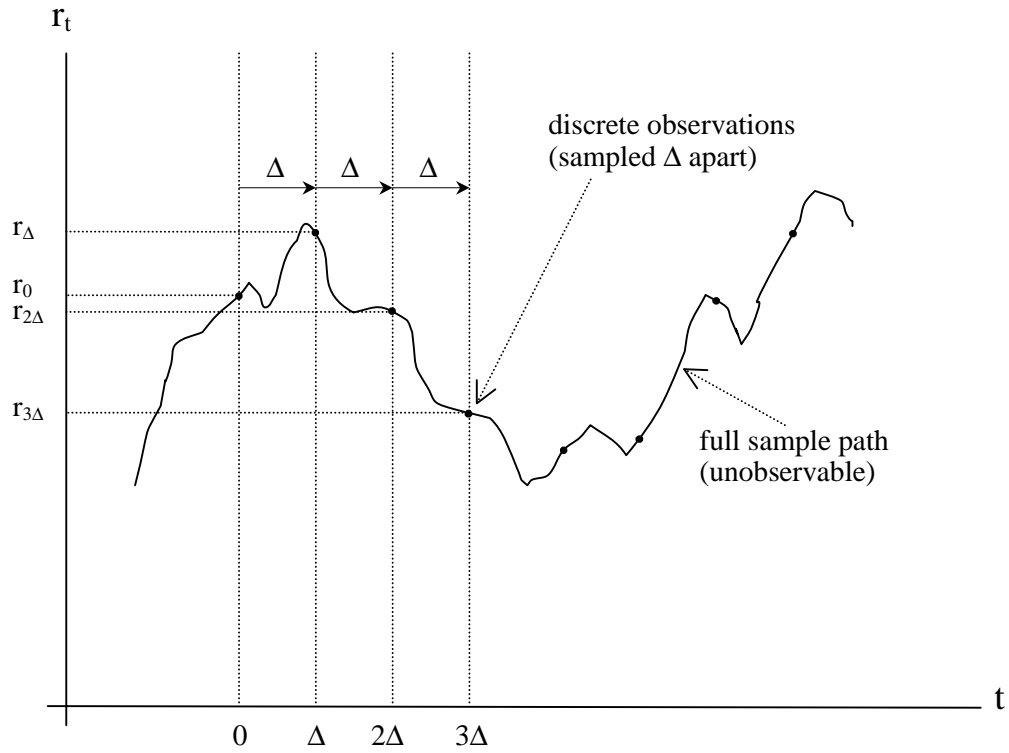
This figure contrasts the discretely-observed sample path from the continuous-time one (which generated the observed sample). It also illustrates the notion of repeated sampling under time-homogeneity, with pairs of successive observations $(r_{i\Delta}, r_{(i+1)\Delta})$, sampled Δ units of time apart and drawn from a common joint distribution which depends on Δ but not i.

Figure 2: Two Sample Paths with No Crossing
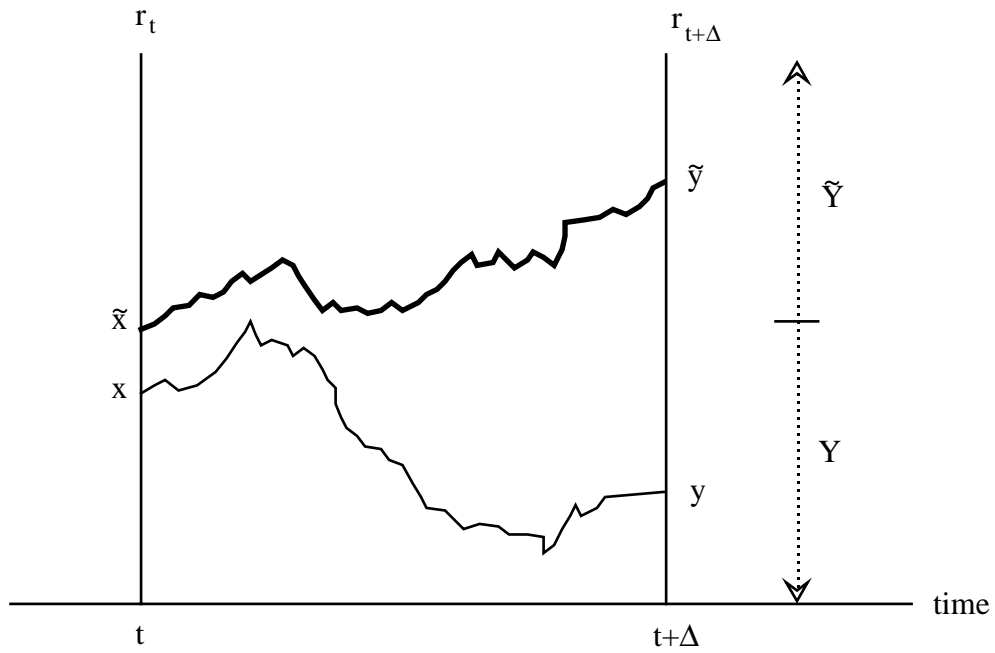

This figure shows the sample paths between two successive sampling dates t and t+Δ followed by the two processes r (thin curve) and $\tilde{r}$ (thick curve), with r positioned at time t at x, below $\tilde{r}$ which starts at $\tilde{x}$. The processes reach time t+Δ having maintained the same order throughout, r finishing in the set Y below $\tilde{r}$ in the set $\tilde{Y}$.

Figure 3: Two Sample Paths Crossing
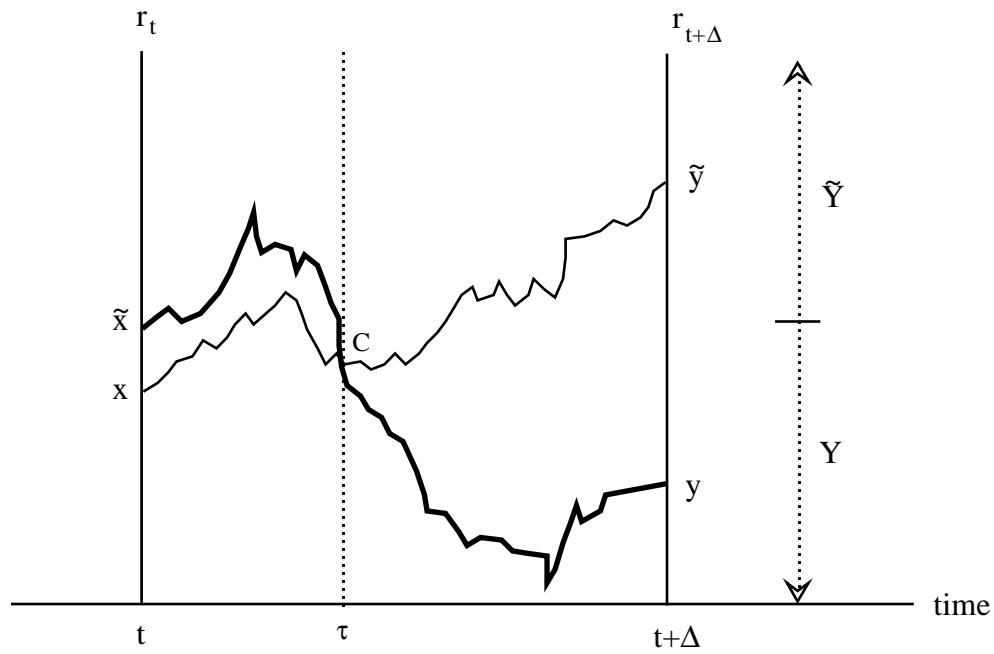
This figure shows the sample paths between two successive sampling dates t and t+Δ followed by the two processes r (thin curve) and $\tilde{r}$ (thick curve), with r starting below $\tilde{r}$ at date t but finishing above it at date t+Δ. If the sample paths are continuous, this can only happen if they cross (at least) once in between. In the figure, the coincidence time (where the two processes are in the same state C) is denoted at τ.

Figure 4: Two Sample Paths with Jumps
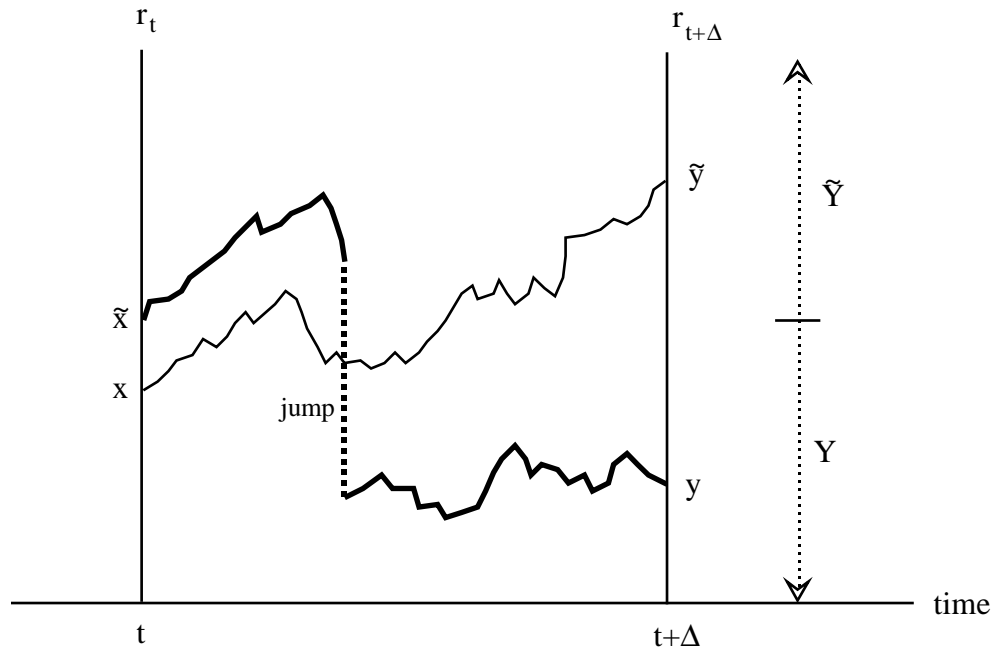
This figure shows the sample paths between two successive sampling dates t and t+Δ followed by the two processes r and $\tilde{r}$, with r starting above $\tilde{r}$ at date t but finishing below it at date t+Δ. By contrast with the situation illustrated in Figure 3, if the sample paths can jump, this reversal of order may happen without the two processes having ever been coincident, i.e., having occupied the same state at the same time. Note that in both Figure 3 and Figure 4, we do not observe what actually happens between t and t+Δ (we only see the process every Δ units of time). We will draw inference about what happened in between from the Δ-apart sample, specifically from features of the transition density p(Δ,y|x).

Figure 5: Discrete Time Markov Chain
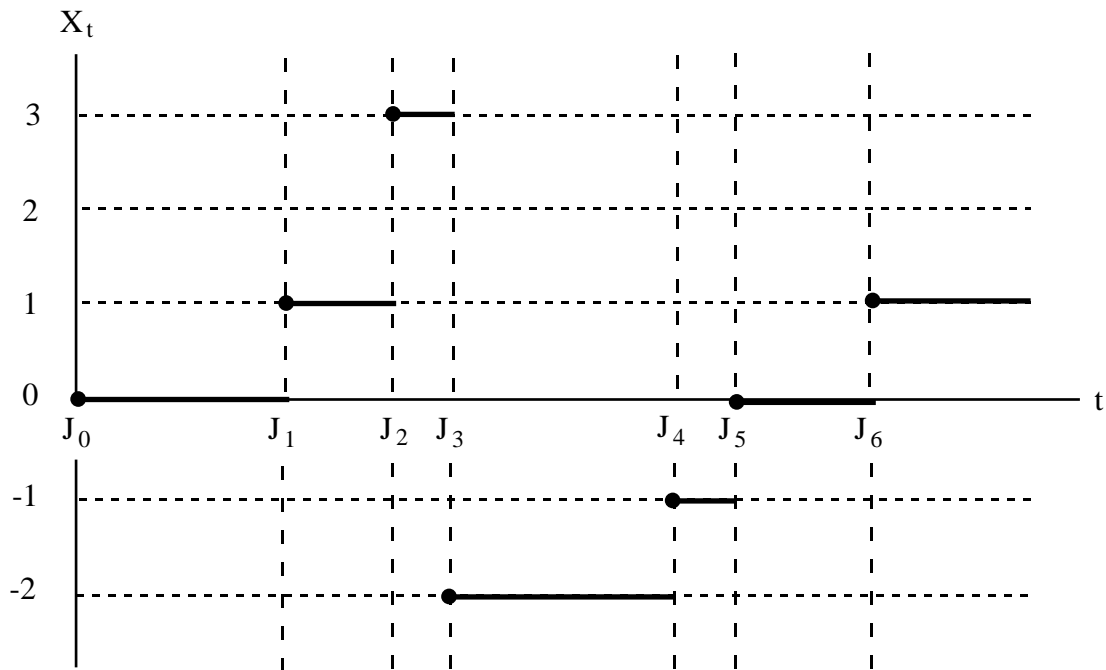
This figure illustrates the concept of continuity of sample paths for a continuous-time, discrete-state Markov chain. By definition, all the state changes are "jumps" and the appropriate notion of continuity distinguishes between "small" or "continuous" jumps, which are those from one state to an immediately adjacent one (+1 or –1), from "large" or "discontinuous" jumps, which are those from one state to a nonadjacent one. The jumps occurring at jump times J_1, J_4, J_5 and J_6 are all of size +1 or –1, so that the process jumps from one state to an immediately adjacent one. By contrast, the jump taking place at jump times J_2 and J_3 are of size +2 and –5 respectively. The first set is compatible with continuity of the sample paths whereas the second set is not.

Figure 6: Discrete-Time, Discrete-State Multinomial Tree
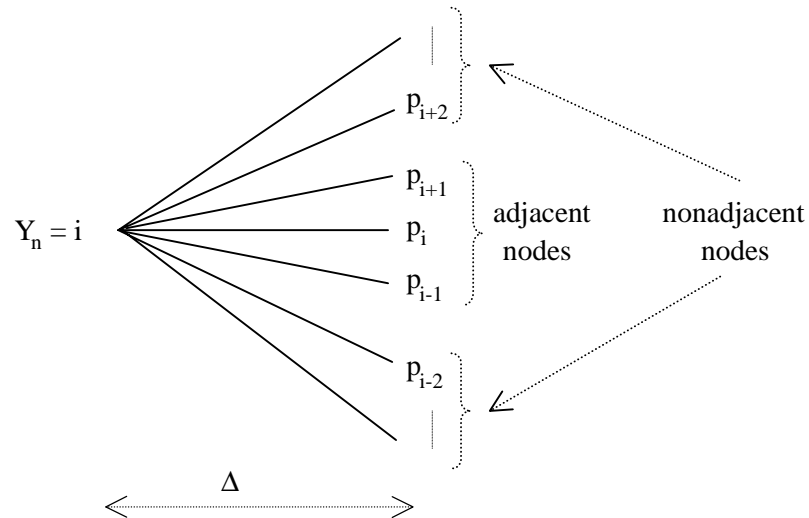
This figure illustrates the concept of continuity of sample paths for a discrete-time, discrete-state multinomial tree. For the approximated continuous-time, continuous-state process to be Markovian, the tree must be recombining. For the approximated process to be a diffusion, the tree must lead to continuous sample paths, which in this case means that all the discrete-time jumps along the tree must occur from one node to an immediately adjacent state. If at date $n\Delta$ the process is in state i, then for the tree to be an approximation to a continuous-path process it must be that the only non-zero branch probabilities are p_{i-1}, p_i and p_{i+1}. Since binomial and trinomial trees only have adjacent nodes, these by construction can only approximate a continuous-time, continuous-state process with continuous sample paths – that is., a diffusion. Conversely, for the tree to approximate a process with discontinuous sample paths, some of the nonadjacent nodes must be attainable.

Figure 7: Diffusion and Jump-Diffusion Transition Functions
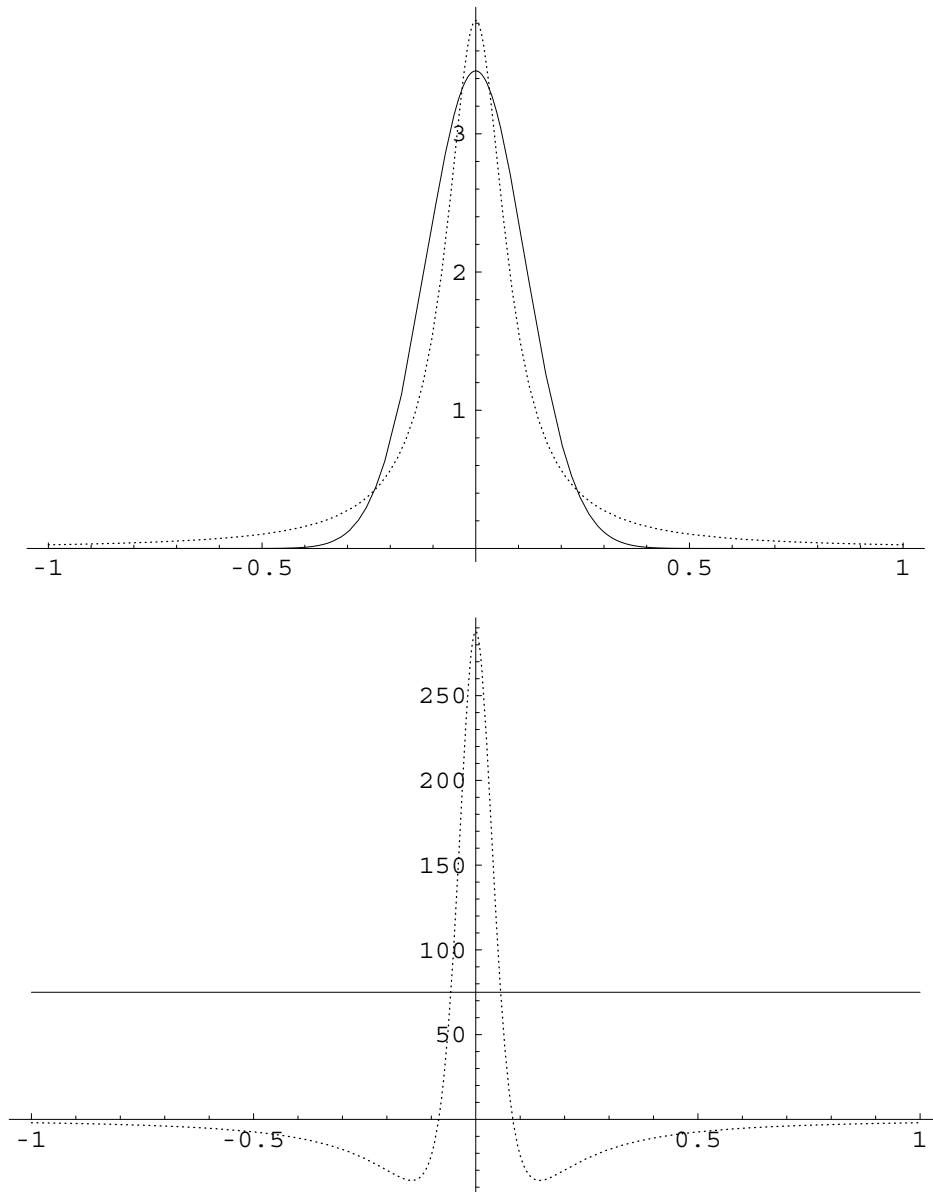
The top plot in this figure represents the transition densities of a Brownian motion (solid curve) and a Cauchy process (dotted curve) as a function of the difference $z = y - x$ between the forward and backward state values. The fact that the tails of the Cauchy distribution are larger is apparent. The bottom plot graphs the criterion function $\partial^2 \text{Ln}(p(\Delta, y \mid x)) / \partial x \partial y$ for both distributions. While the criterion is always positive for the Brownian motion (solid curve), it is negative in the tails (where z is large in absolute value) for the Cauchy distribution (dashed curve).

Figure 8: SPX Implied Volatility Smile and State-Price Density
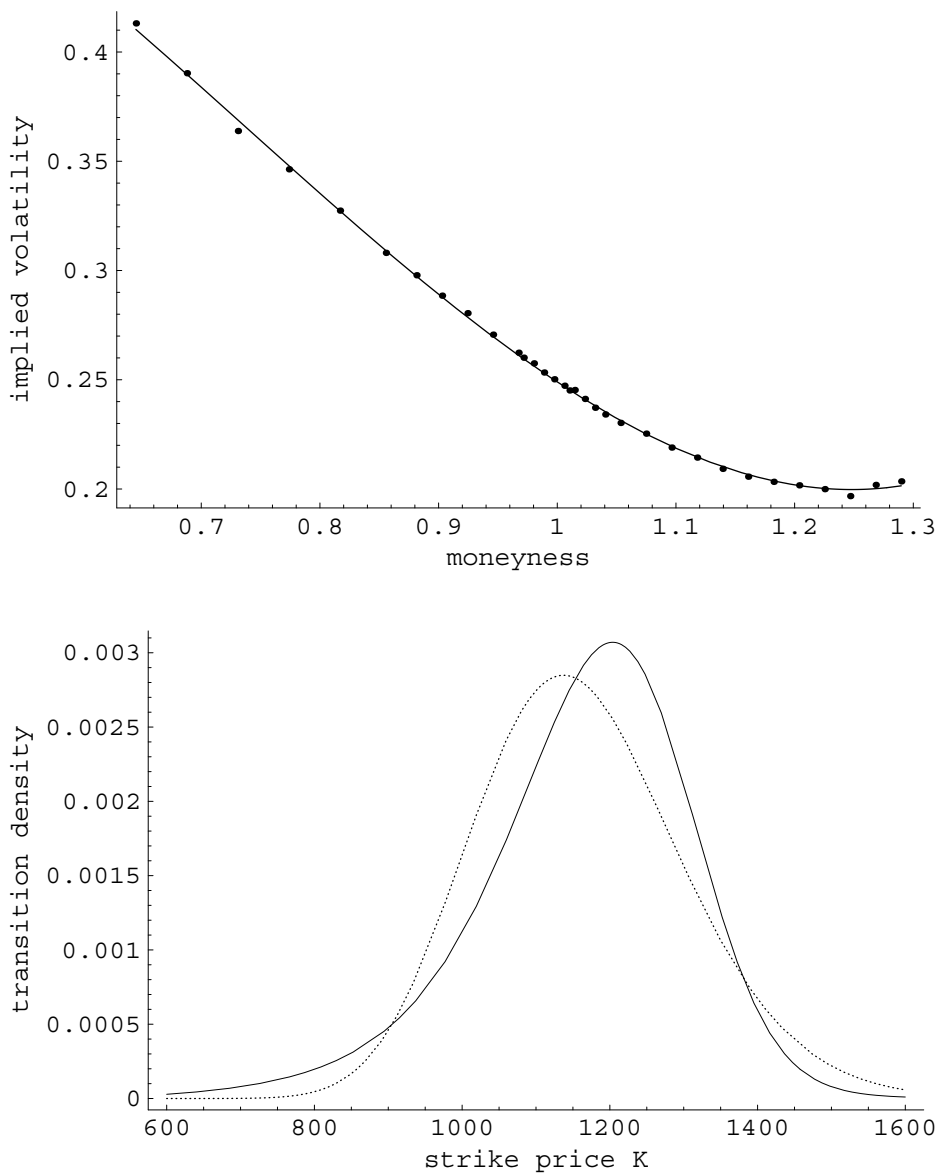
The top plot in this figure reports the fitted implied volatility function $K/F_\Delta \mapsto \sigma_{IMP}(K/F_\Delta)$. The dots represent the actual implied volatilities from Table I. The parameter estimates are reported in Table II. The solid curve bottom plot in the figure represents the implied transition density $K \mapsto p(\Delta, K \mid x_0)$ for the maturity Δ and current index value x_0 described in Table I. For comparison purposes, the Black-Scholes state-price density (dotted curve) evaluated at the at-the-money implied volatility is also included. The skewness of the implied density is apparent. All by itself, however, skewness is not sufficient indication of non-diffusion behavior.

Figure 9: Diffusion Criterion Applied to the State-Price Density
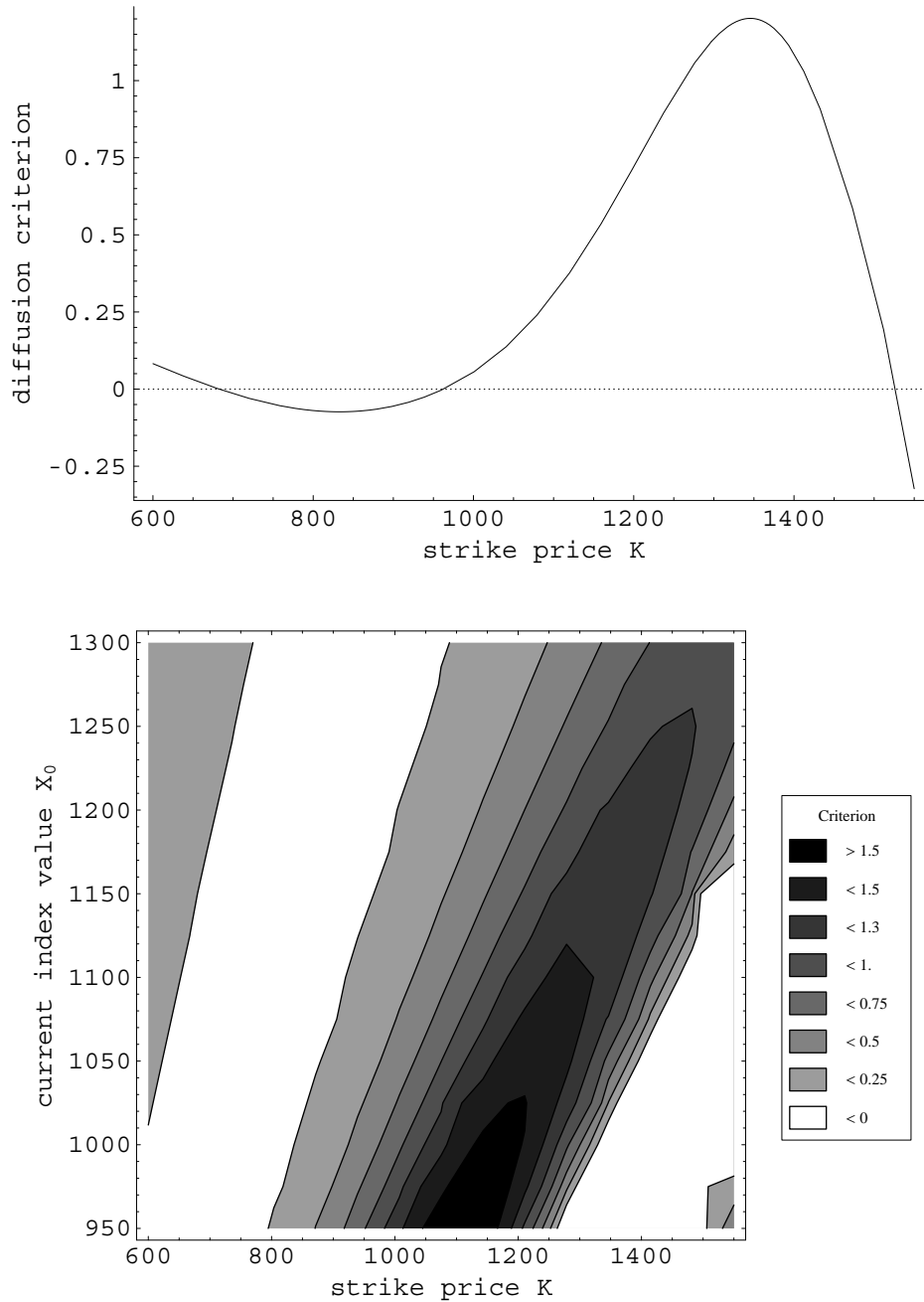
The two plots in this figure determine whether the option-implied transition density is compatible with a diffusion model for the underlying (risk-neutral) asset price dynamics. The top plot graphs the criterion function $K \mapsto \partial^2 Ln(p(\Delta, K \mid x_0)) / \partial x_0 \partial K$ for the fixed Δ and x_0 corresponding to the option data that are given in Table I. For display purposes, the criterion function is multiplied by a fixed constant (this is irrelevant to the conclusions, since we only care about the sign of the criterion). Notice that there are regions where the criterion function becomes negative The bottom plot is a contour plot, for the fixed Δ in the data, of the three-dimensional surface $(K, x_0) \mapsto \partial^2 Ln(p(\Delta, K \mid x_0)) / \partial x_0 \partial K$. White areas in the contour plot indicate regions where the criterion function is negative. Anything short of nonnegativity of the criterion function for all values of (K, x_0) is incompatible with an underlying diffusion model. The conclusion from the analysis is that there exists no diffusion model that could have generated the transition density (reported in Figure 8) which prices these options.