Simple Option Pricing and the Leverage Effect*

Stefan De Wachter†
Department of Economics, University of Oxford

December 22, 2004

Abstract

This paper develops a nonparametric test to investigate whether any of two classes of "simple" models can rationalize observed option prices. These classes are (i) processes with independent returns and (ii) univariate Markov processes. The practice of recalibration is mimicked by only imposing minimal stability requirements on the candidate pricing models. The main finding is that processes with independent returns are inadequate even for the purpose of fitting the cross-section and term-structure on a single day. Univariate Markovian processes, however, perform much better. This is due to their ability to capture leverage.

Keywords: option pricing, martingale measure, Markov property, leverage
JEL classification: G13, C12, C52

1 Introduction

The development of option pricing methods since Black and Scholes' seminal 1973 article has primarily focussed on exploring alternative assumptions for the dynamics of the underlying asset. Two different strategies to this end co-exist. One strand of the academic literature has followed the evolution in the econometric literature towards models featuring increasingly complex serial dependence (in particular ARCH (Engle (1982), Bollerslev (1986)) and stochastic volatility (SV) (Taylor (1986))). Examples are Hull and White (1987), Stein and Stein (1991), Heston (1993a), Heston and Nandi (2000), Nicolato and Venardos (2003) and Carr and Wu (2004).

A second strand of the literature retains its focus on serial dependence structures similar or slightly more general than "Black-Scholes" geometric Brownian motion. In addition to early papers such as Merton (1976) or Cox and Ross (1976), recent contributions by Madan and Seneta (1990), Rubinstein (1994), Madan, Carr, and Chang (1998), Kou and Wang (2001), Carr and Wu (2003), Carr, Madan, Geman, and Yor (2005) and several others

*This paper is based on part of my PhD dissertation at the University of Oxford. I would like to thank Richard Spady for excellent research supervision and Steve Satchell, Hyun Shin and participants at the Nuffield College econometrics seminar for many constructive comments.

†Contact address: Nuffield College, New Road, Oxford OX1 1NF, UK; email stefan.dewachter@economics.ox.ac.uk.

cited below consider option pricing on processes with a simple dependence structure, using increasingly flexible parametrizations.

While this by itself might suffice to justify further study of such simple models, it is their popularity with practicioners that provided the initial motivation for the present work. Indeed, the technical machinery supporting option trading decisions in practice tends to be based on "tweaked" versions of the Black-Scholes model rather than on the most complex methods from the academic literature. The new proposals for pricing models based on processes with a simple dependence structure can in that sense be seen as a potential improvement over currently popular pricing tools without sacrificing much of their computational and conceptual simplicity. The extent of their misspecification is likely to be larger than that of the more complex models, but this problem is commonly circumvented by frequent recalibration.

In this paper, I develop and apply a device for analyzing the potential of simple option pricing models. This device is able to discriminate between "simple" and "complex" models based on very short observation periods. Hence it is suited for checking whether simple models are sufficiently flexible to price options whilst remaining stable over short periods. This is in line with the common practice in the market of using simple models for pricing and hedging, and resorting to fairly frequent recalibration to deal with potential misspecification. Of course, recalibration is inconsistent with the theory behind the models; what I want to discover is not whether simple models reflect the true DGP behind several decades of data - they certainly do not, nor does any other model investigated to date - but rather how sophisticated a pricing model needs to be in order to be a potentially useful improvement over tools currently used in practice.

The method consists of two tests, both based on a consistency relationship between risk-neutral transition densities[1]. The first holds for Markov processes and is known as the Kolmogorov-Chapman equation; the second is a specialization thereof for processes with independent returns. Because the tests are based on a property of an entire class of processes rather than on a particular specification, they are nonparametric in nature. This allows me to make statements about the practical relevance of the entire class without having to specify a parametric model. The drawback is that the procedure does not produce a fully specified "best" model; however, it does generate an estimate of an implied forward density: the density, implied by current and recent option prices, of the underlying at some distant point in the future given its value at an intermediate point in the future.

Incidentally, the dependence of this forward density on its conditioning variable plays a central role in the interpretation of my main result. Indeed, the main discovery is that when stability over a single week or less is imposed, evidence against univariate Markovian models is weak whereas models with independent returns are clearly inadequate - a result which is at odds with several statements in the literature. In fact, models with independent returns cannot even explain the cross-section and term-structure of options observed on a single day! The root cause of this performance differential is the inability of processes

[1] Also called state-price densities (SPDs) in some of the literature, these are conditional densities of the price of the underlying asset at the options' maturity date given current information.

with independent returns to capture asymmetric volatility[2] (leverage) as exemplified by the negative dependence of the dispersion of the Markov-implied forward density on the value of the underlying asset at the start of the forward period. When longer periods of stability are required, univariate Markov processes are rejected more convincingly by the data, although no easily interpretable mispricing pattern emerges.

This paper is organized as follows. I start by reviewing the most closely related literature in Section 2. Section 3 defines precisely what is meant by "simple models", lists several examples, and introduces the test. Section 4 contains an application to S&P 500 options data, and Section 5 concludes. The Appendix contains a derivation of the asymptotic distribution for the test.

2 Related literature

A vast number of studies have addressed the comparative performance of parametric option pricing models. The overall conclusion is that even the most complex models are not able to explain observed price fluctuations accurately enough, especially - but not exclusively - when model stability over long periods of time is required (see e.g. Bates (2000) or Bakshi, Cao, and Chen (1997)). I refer the interested reader to the recent survey by Bates (2003) and focus here on the two contributions that are closest in spirit to the present paper.

The most recent of these is Buraschi and Jackwerth (2001) [BJ], who develop a number of time-series-based tests for a class of deterministic volatility (DV) models based on spanning arguments. Their methods aims to test whether all contingent claims can be perfectly hedged using only the riskless asset and the underlying. They find that two assets are not sufficient to span the observed payoffs and hence that DV models that imply market completeness are not general enough as a description of risk-neutral dynamics over a period of several years.

The important differences between my approach and that of Buraschi and Jackwerth (2001) are twofold. Firstly, my method is based on very small samples and can thus evaluate model stability over recalibration horizons comparable to those used in practice whereas BJ's approach requires several years of data. Secondly, the content of the null hypothesis differs in that I allow for models that imply market incompleteness such as jump-diffusions.

A second attempt to gauge the adequacy of a certain type of DV models is presented by Dumas, Fleming, and Whaley (1998) [DFW]. They parametrize the local volatility function in a generalized Brownian motion[3] using 4 different specifications. Each of these is calibrated to a number of options observed at an initial date. In-sample fit and out-of-sample performance on prices observed 1 week later are then examined. DFW report a serious lack of stability of their DV models; in addition, increasing the flexibility of the model too much has adverse effects on its out-of-sample performance.

My test is related to DFW's approach in that it uses a dataset covering a short stretch of

[2] I point out that this is different from skewness in conditional distributions or, equivalently, negatively sloped implied volatility curves. Indeed, this feature *can* be captured by processes with independent returns.

[3] A generalized Brownian Motion is the solution to the stochastic differential equation $dS = rSdt + \sigma(S,t)dW$. $\sigma(S,t)$ is called the local volatility function. See also Cox and Ross (1976) and Dupire (1994).

time and works only under the martingale measure[4]. However, my null hypothesis comprises a considerably more general class of processes than DFW's parametrized DV models.

3 Construction of the tests

3.1 Notation and definitions

As mentioned in the introduction, a "simple model" is defined to be a model that is either "univariate Markovian" or has "independent returns". This subsection defines those concepts and some of their implications. The next subsection contains examples illustrating which kinds of well-known models do and do not share these properties.

In what follows, I use the familiar notation of a stochastic process $\{S_t\}$, the underlying asset, on a filtered probability space $(\Omega, Q, I, \{I_t\})$. The probability measure Q is the equivalent martingale measure. The physical probability measure does not feature in this paper. State-price densities (SPDs) are transition densities, with respect to Q, of S_T given information available at time t ($t < T$), and are denoted by $q_{t,T}(S_T|I_t)$[5]. I will use the notational convention that the subscript of q contains the time-interval, and the arguments indicate the scale[6] and what is being conditioned on. The sigma-field generated by S_t (information contained in S_t) is denoted by $I(S_t)$.

Definition 1 *The stochastic process $\{S_\tau\}$ satisfies the **Markov property** iff for all t, T ($t < T$)*

$$q_{t,T}(S_T|I_t) = q_{t,T}(S_T|I(S_t))$$

This essentially means that once S_t is known, there is nothing else to learn at time t (from the past or present) about the distribution of S_T. S can be vector-valued in this formulation; I will call a scalar-valued process $\{S_t\}$ satisfying the above definition a **univariate Markov process**. In the following I will simplify notation by writing $q_{t,T}(S_T|I(S_t))$ as $q_{t,T}(S_T|S_t)$.

The test will be based on a property of Markov processes, known as the Kolmogorov-Chapman equation. Using the notation introduced before, it is formulated in

Corollary 1 *(The Kolmogorov-Chapman equation) If the process $\{S_t\}$ satisfies the Markov-property, then one has the following relationship between transition densities:*

$$q_{t,T_2}(S_{T_2}|I_t) = \int_{S_{T_1}} q_{T_1,T_2}(S_{T_2}|S_{T_1})\ q_{t,T_1}(S_{T_1}|I_t) dS_{T_1} \qquad (1)$$

for all $t < T_1 < T_2$.

[4] BJ, in contrast, essentially replace theoretical pricing relationships by sample averages, explaining their need for a large amount of data.

[5] The term state-price density is also used to refer to a different object, namely the density (Radon-Nikodym derivative) of the pricing measure Q with respect to the physical probability measure. Here I use SPD as a synonym for risk-neutral transition density.

[6] For example the density of $\log(S_T)$ is denoted by $q_{t,T}(\log(S_T))$. This is a slight abuse of notation.

Equation (1) states that the market's opinion at t about the risk-adjusted conditional density of S_{T_2} given S_{T_1} (where S_{T_1} is of course not known at t) depends **only** on S_{T_1}, and not on, for instance, the price-*path* from t to T_1 or any other variable at any point in time.

A particular class of univariate Markov processes are processes with *independent returns*. When this property holds, returns for two non-overlapping time periods are independent random variables, or otherwise stated, that the logarithm of the process is a process with independent increments. Translated in terms of transition densities, one writes

Definition 2 *A (scalar-valued) stochastic process* $\{S_\tau\}$ *is said to have* ***independent returns*** *iff it is has the univariate Markov property and, in addition, for all* t, Δ, *the density of* $\log(\frac{S_{t+\Delta}}{S_t})$, *denoted by* $q_{t,t+\Delta}(\log(\frac{S_{t+\Delta}}{S_t}))$, *satisfies*

$$\forall t, \Delta : q_{t,t+\Delta}(\log(\frac{S_{t+\Delta}}{S_t})|S_t) = q_{t,t+\Delta}(\log(\frac{S_{t+\Delta}}{S_t}))$$

This implies that if $t' \geq t + \Delta$, for a given x

$$q_{t,t+\Delta}(\log(x)) - q_{t',t'+\Delta}(\log(x)) = m_x(t, t')$$

for some deterministic function m_x. In other words, the distribution of the log-return is always the same up to a deterministic function of time[7] for a given time-interval Δ and does therefore not depend on the price-level at the start of the interval.

The Kolmogorov-Chapman equation is obviously still valid, but the integral in (1) may in some cases be easier to compute. Indeed, under independent returns, equation (1) reduces to

$$q_{t,T_2}(\log(\frac{S_{T_2}}{S_t})) = \int_{S_{T_1}} q_{T_1,T_2}(\log(\frac{S_{T_2}}{S_{T_1}}))\ q_{t,T_1}(\log(\frac{S_{T_1}}{S_t}))dS_{T_1} = q_{T_1,T_2} * q_{t,T_1} \qquad (2)$$

where $*$ denotes convolution[8].

3.2 Examples

I now list some well-known option pricing models based on processes with the above properties. The most famous example of a process with **independent returns** is "Black-Scholes" geometric Brownian motion: with r denoting the riskless interest rate, which is assumed to be deterministic and constant, one can write

$$\begin{aligned} d\log S_t &= (r - \sigma^2/2)dt + \sigma dW \\ \log(S_{T_1}) - \log(S_t) &= const + \sigma(W_{T_1} - W_t) \\ \log(S_{T_2}) - \log(S_{T_1}) &= const + \sigma(W_{T_2} - W_{T_1}) \end{aligned}$$

The right-hand sides of the two last expressions are clearly independent. A related example is the "binomial tree" implementation of the Black-Scholes model of Cox, Ross, and Rubin-

[7] In most models of this type, $m = 0$. However, time-homogeneity is not required and not implied by the concept of independent returns, nor by the univariate Markov property.

[8] Of course, this only simplifies the computation if the characteristic functions of the SPDs are easily calculable.

stein (1979), as are further discrete-state and -time variations on the idea such as trinomial trees with constant volatility as well as basic extensions of the Black-Scholes model where deterministic calendar-time dependent volatility is introduced (see e.g. Rebonato (2000)).

When building continuous-time models with independent returns, Brownian motion is not the only possible building block. Indeed, any Lévy process can be used, or any combination of Lévy processes. The most well-known example of this kind is Merton's (1976) jump process, according to which the process evolves, under the martingale measure, according to the equation

$$dS_t = (r - \lambda m)S_t dt + \sigma S_t dW_t + S_t(J_t - 1)d\pi_t \tag{3}$$

where W is Brownian motion, J_t are serially independent positive random variables with mean m with a distribution to be specified by the modeller, and π_t is a Poisson process with intensity λ. Additionally, the final term in this equation could be replaced with another Lévy process, even though this may further complicate computation of option prices and simulation. Examples of this jump-diffusion or pure jump-process approach can be found, amongst others, in Madan and Seneta (1990), Madan and Milne (1991), Ahn (1992), Amin (1993), Heston (1993b), McCullogh (1996), Eberlein, Keller, and Prause (1998), Madan, Carr, and Chang (1998), Kou and Wang (2001)[9], Carr, Geman, Madan, and Yor (2002), Kou (2002), Carr and Wu (2003) and Cartea and Howison (2004).

All processes with independent returns are also **univariate Markovian**. The latter class is much larger, and contains the following examples of processes that have been used as the basis for option pricing models. Probably the best known univariate Markovian model that does not have independent returns is so-called generalized Brownian motion:

$$dS_t = rS_t dt + \sigma(S_t, t)\, dW_t \tag{4}$$

for some choice of the function $\sigma(S_t, t)$. The first occurrence of a process of this type, with the choice $\sigma(S_t, t) = \sigma S_t^\rho$, was proposed in Cox and Ross (1976) (see also Beckers (1980) and Delbaen and Shirakawa (2003)) and is known under the name "constant elasticity of variance (CEV) process". In early applications, the choice $\rho = 1/2$ was most common. Subsequently, a number of studies have focussed on flexible calibration of the volatility function rather than the tight parametrization of the CEV process - see Dupire (1994) and Dumas, Fleming, and Whaley (1998) (discussed above) for continuous-time examples and Rubinstein (1994) for a flexible binomial tree implementation of the same idea. Model (4) is extremely popular among practicioners as it is conceptually not too different from the standard Black-Scholes model - it also implies market completeness - yet allows to capture both the smile and the term-structure on any (single) given day - see e.g. Jackson, Suli, and Howison (1999) for details regarding implementation. Carr, Madan, Geman, and Yor (2005) present a treatment of models of type (4) where the Brownian motion W_t is replaced by some other Lévy process.

Combining a model of the type of equation (4) with different kinds of Lévy processes, possibly dependent on the contemporaneous price level, is a logical extension. Processes

[9] See also http://www.thi.com/seminars/apr132000.shtml for information on a seminar organized by the quant firm TechHackers on this topic.

of this type have remained relatively unexplored, but see Henderson and Hobson (2003) for some theoretical considerations and Andersen and Andreasen (2000) for an empirical example.

The history of option pricing theory has seen a considerable amount of models that are **not univariate Markovian**. These are based on either GARCH processes (in which conditional volatility depends on past returns) or stochastic volatility processes (which are bivariate, but not univariate, Markov processes), possibly enhanced with Poisson (or more general) jumps. References to a small selection of well-known contributions in this area were given in the introduction to the paper.

3.3 Testing principle

The aim of this subsection is to construct statistical tests of the null hypotheses that the risk-neutral dynamics of the underlying are univariate Markovian / have independent returns against the alternative that they are governed by some further unspecified martingale measure. These tests will be based on equations (1) and (2), respectively. To show how these can be exploited to construct a test, Figure 1 presents a typical situation encountered in option markets.

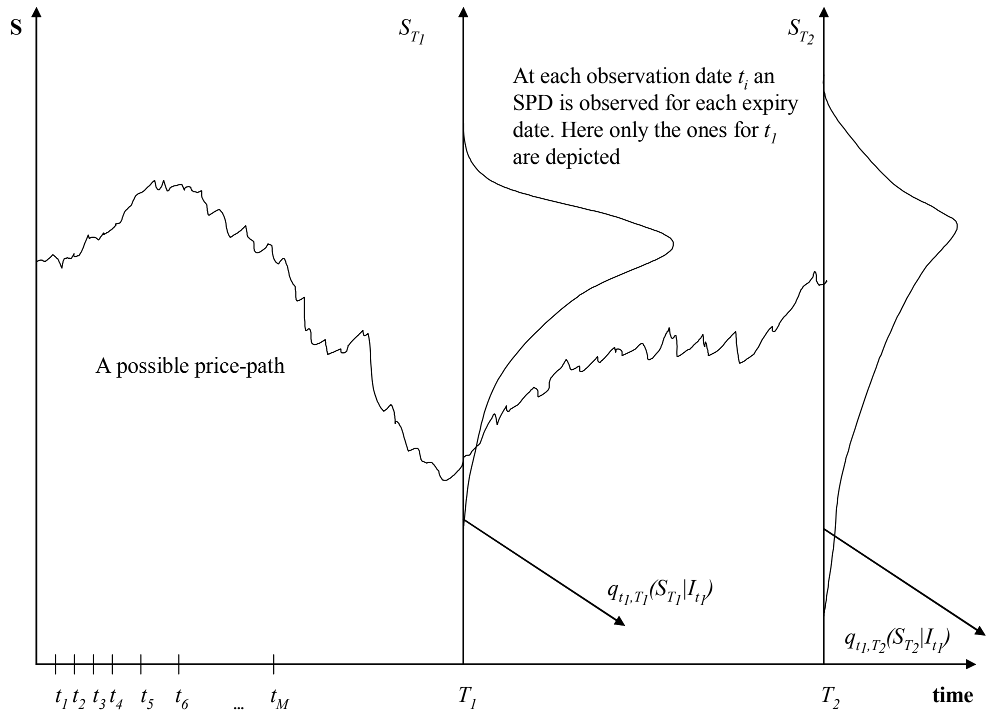

Figure 1: Observable risk-neutral transition densities (SPDs) in an options market. The SPDs are drawn in the "third dimension" as indicated by the axes coming "out of the paper".

A number of European-style options written on an asset with price S and expiring at fixed dates T_1 or T_2 are traded in a liquid market at all dates t_j $(j = 1, ..., M)$. Denote by $C_{t,T}^K$ the time-t price of a call option with strike price K expiring at T. Assuming that the interest rate r and the dividend rate d are deterministic functions of time, the exposition can be simplified by setting $r = d = 0$. In the absence of arbitrage, the price at t_j of an option expiring at T_e $(e = 1, 2)$ is then given by the expectation of its payoff function with respect to the SPD $q_{t_j,T_e}(S_{T_e}|I_{t_j})$.

Now assume for the sake of the argument that observing a cross-section[10] of option prices is "equivalent" to observing the corresponding SPD. That is, assume that each $q_{t_j,T_e}(S_{T_e}|I_{t_j})$ is known. Since the time-interval between T_1 and T_2 remains fixed (because expiration dates are fixed), equation (1) applies at each t_j if the risk-neutral process has the univariate Markov property. As time t goes by and the transition densities $q_{t_j,T_e}(S_{T_e}|I_{t_j})$ are observed, one obtains the following system of M equations, which constitutes the basis for the test

$$\begin{aligned} q_{t_1,T_2}(S_{T_2}|I_{t_1}) &= \int_{S_{T_1}} q_{T_1,T_2}(S_{T_2}|S_{T_1})\; q_{t_1,T_1}(S_{T_1}|I_{t_1})\, dS_{T_1} \\ &\vdots \\ q_{t_M,T_2}(S_{T_2}|I_{t_M}) &= \int_{S_{T_1}} q_{T_1,T_2}(S_{T_2}|S_{T_1})\; q_{t_M,T_1}(S_{T_1}|I_{t_M})\, dS_{T_1} \end{aligned} \tag{5}$$

Note that there is only one unobserved SPD, $q_{T_1,T_2}(S_{T_2}|S_{T_1})$, and this SPD appears in each equation. All other SPDs in the system are observed. If the underlying asset is not univariate Markovian, there may still exist a density function $\widetilde{q}_{T_1,T_2}(S_{T_2}|S_{T_1})$ that makes any given equation in (5) hold separately, but this function will generally not be constant across equations, i.e. as time t_j goes by. For the case of independent returns, a similar system based on equation (2) holds.

The idea of the tests is to check whether there exists a function $q_{T_1,T_2}(S_{T_2}|S_{T_1})$ that makes system (5) hold, or equivalently, whether imposing system (5) results in a significant deterioration of pricing performance. Details are discussed in the next subsection.

One may remark that the system of Kolmogorov-Chapman equations (5) does not capture all restrictions on the data implied by independent returns / the Markov property. Indeed, no constraints are placed on the relationship between the T_1 SPDs $q_{t_j,T_1}(S_{T_1}|I_{t_j})$; equations (5) focus exclusively on the "forward" properties in the data. Nevertheless, a test based on (5) will be able to discriminate against most interesting alternatives. One may expect that the test for independent returns (based on system (5) with equations of the type (2)) will detect stochastic volatility as long as the mean reversion of volatility is slow enough relative to the options' time-to-maturity[11]. It will also be very powerful against alternatives with asymmetric volatility ("leverage"). The test for the univariate Markov property (equations of type (1)) will have power against models with slowly reverting stochastic volatility, *unless* this volatility is very strongly asymmetric, i.e. strongly correlated with the underlying asset. These insights hold irrespective of whether the underlying processes have continuous price-paths or not.

[10] "Cross-section" refers to contracts observed on the same day, with common time-to-maturity, but with different strike prices.

[11] If volatility reverts quickly, only $q_{t_j,T_1}(S_{T_1}|I_{t_j})$ will be affected by fluctuations in current volatility. No relationship between $q_{t_j,T_1}(S_{T_1}|I_{t_j})$ is imposed by system (5).

3.4 The test-statistic and its distribution

In order to subject the relationships (5) to empirical scrutinity, two problems need to be addressed. Firstly, as equation (6) below shows, system (5) implies a restriction on the pricing function for T_2 options, relating this function to the T_1 SPD. Although this is a deterministic relationship, actual observed option prices are unlikely to correspond exactly to *any* pricing function. Indeed the presence of bid-ask spreads, price asynchronicities and staleness, price discreteness, and even plain mispricing necessitates the inclusion of a "measurement error term" in the representation of option prices. When empirically verifying whether call prices obey the relationships in expression (6) below, one needs to account for measurement error both in T_2 option prices $C^K_{t_j,T_2}$ and in the SPDs $q_{t_j,T_1}(S_{T_1}|I_{t_j})$, which need to be extracted from T_1 options. While no strict consensus exists in the literature about the choice of a statistical model for option pricing errors[12], I will follow other authors (e.g. Engle and Mustafa (1992) or Yatchew and Härdle (2003)) by representing these errors additively as independent (across both time and cross-section) random variables.

The second problem concerns the specification of the state-price densities. Whereas the hypothesis at hand is strictly a nonparametric one - apart from the martingale property, economic theory does not pose any restrictions on the state-price densities - the typically small sample size (about 20 to 60 options per observation day) leads me to prefer a flexible parametric approach[13]. While sensible from a practical point of view, this choice does lead to a number of problems when deriving an asymptotic distribution theory for the test. In particular, because my parametrization for state-price densities is not closed under convolution (and, a fortiori, under the operation defined by the Kolmogorov-Chapman equation), the initial nonparametric nested testing problem is transformed in a parametric nonnested one. Further details on this issue can be found in the Appendix.

Construction of the test-statistic proceeds as follows. Multiplying each equation by $\max(S_{T_2} - K, 0)$ and integrating over S_{T_2}, one can rewrite system (5) as

$$C^{K_{j,i}}_{t_j,T_2} = \int_{S_{T_1}} C_{T_1,T_2}(S_{T_1}, K_{j,i})\; q_{t_j,T_1}(S_{T_1}|I_{t_j})\, dS_{T_1} \quad j = 1, ..., M\;;\quad i = 1, ..., n_{2,j} \qquad (6)$$

where n_{2j} is the number of T_2-options observed at t_j. To arrive at the system of equations (6) from (5), one uses the fact that $\int_{S_{T_2}} \max(S_{T_2} - K, 0)\; q_{t_j,T_2}(S_{T_2}|I_{t_j})\; dS_{T_2} =: C^K_{t_j,T_2}$ and defines the forward-period call pricing function $C_{T_1,T_2}(S_{T_1}, K) := \int_{S_{T_2}} \max(S_{T_2} - K, 0)\; q_{T_1,T_2}(S_{T_2}|S_{T_1})\, dS_{T_2}$. Equations (6) express T_2 option prices using the restrictions imposed by system (5). They do not add any additional information beyond that contained in system (5).

Referring back to Figure 1, one sees that at each date t_j $(j = 1, ..., M)$, $n_{e,i}$ option prices are observed that expire at T_e $(e = 1, 2)$. For each e and j, stack these $n_{e,j}$ observed prices in the vector C_{t_j,T_e}, the strike prices in the vector K_{t_j,T_e}, and the corresponding (unobserved) measurement errors in the vector ε_{t_j,T_e} $(e = 1, 2)$. For each observation date t_j, each of the two available cross-sections is associated with a separate SPD. Let θ_{t_j,T_e} denote the

[12] See Clément, Gouriéroux, and Monfort (2000) and Jacquier and Jarrow (2000) for related comments.
[13] Details on the functional form used can be found in the Appendix.

parameter vector corresponding to the parametric SPD $q_{t_j,T_e}(S_{T_e}|I_{t_j})$. For options observed at t_j expiring at T_e $(e = 1, 2)$ the pricing functions are grouped in the $n_{e,j}-$vector of functions $C_{t_j,T_e}(K_{t_j,T_e};\theta_{t_j,T_e})$ (one component for each strike price). Each cross-section can now be written as:

$$C_{t_j,T_e} = C_{t_j,T_e}(K_{t_j,T_e};\theta_{t_j,T_e}) + \varepsilon_{t_j,T_e} \qquad j = 1, ..., M \quad e = 1, 2 \tag{7}$$

All parameter vectors θ_{t_j,T_e} can be estimated by nonlinear least squares. Note that this can be done cross-section by cross-section because of the assumption that measurement errors are independent over time and across maturities. For example, if there are 5 observation days and 20 options per maturity on each day, one estimates 10 SPDs based on 20 observations each. The resulting estimates of the functions $C_{t_j,T_e}(K_{t_j,T_e};\theta_{t_j,T_e})$ and the corresponding estimates of $q_{t_j,T_e}(S_{T_e}|I_{t_j})$ are consistent for the "true" call-pricing functions resp. SPDs up to misspecification error. This holds regardless of whether the martingale measure satisfies the univariate Markov property / independent returns.

If the risk-neutral process driving the underlying asset satisfies independent returns, t_j-prices of options expiring at T_2 can also be expressed in terms of an SPD constructed from $q_{t_j,T_1}(S_{T_1}|I_{t_j})$ and $q_{T_1,T_2}(S_{T_2}|S_{T_1})$ - this construction is given by expression (2). Denoting by θ_{Ind} the parameter-vector of $q_{T_1,T_2}(S_{T_2}|S_{T_1})$ and stacking the resulting pricing functions for T_2 options as above one obtains

$$C_{t_j,T_2} = C^{Ind}_{t_j,T_2}(K_{t_j,T_2};\theta_{t_j,T_1},\theta_{Ind}) + \varepsilon^{Ind}_{t_j,T_2} \quad j = 1, ..., M \tag{8}$$

where $C^{Ind}_{t_j,T_2}(K_{t_j,T_2};\theta_{t_j,T_1},\theta_{Ind})$ corresponds to the right-hand side of equation (6). Estimation of all parameters can now not be done separately because θ_{Ind} is constant across all observation days. To keep the estimation procedure numerically tractable, one can first estimate θ_{t_j,T_1} separately as before, and then calculate $\widehat{\theta}_{Ind}$ by minimizing squared pricing errors in

$$C_{t_j,T_2} = C^{Ind}_{t_j,T_2}(K_{t_j,T_2};\widehat{\theta}_{t_j,T_1},\theta_{Ind}) + \varepsilon^{Ind}_{t_j,T_2} \quad j = 1, ..., M \tag{9}$$

for all $j = 1, ..., M$ simultaneously. The test-statistic V is based on a comparison of pricing errors of T_2-options obtained under the null (i.e. via equation (8)) and those obtained under the alternative (via equation (7)), in particular

$$V = \sum_{j=1}^{M}\left(\widehat{\varepsilon}^{Ind\prime}_{t_j,T_2}\widehat{\varepsilon}^{Ind}_{t_j,T_2} - \widehat{\varepsilon}^{\prime}_{t_j,T_2}\widehat{\varepsilon}_{t_j,T_2}\right) \tag{10}$$

The same procedure can be followed for the Markov test.

The derivation of a distribution theory for the statistic (10) is somewhat involved; the reader is referred to the Appendix (Section 6.2) for details and further motivation. The main result is the following:

Proposition 1 *Under the null hypothesis that the risk-neutral process rationalizing equilibrium option prices has independent returns / the univariate Markov property and under* *appropriate conditions on the distribution of the measurement errors* ε *one has that*

$$V - \left\{ \sum_{j=1}^{M} n_{2j}\Delta + \sum_{j=1}^{M} \frac{\sqrt{n_{2j}}}{\sqrt{n_{1j}}} \sqrt{n_{2j}}\mu_j' \sqrt{n_1}(\widehat{\theta}_{t_j,T_1} - \theta^*_{t_j,T_1}) \right\}$$

is approximately distributed as a weighted sum of independent $\chi^2(1)$ *distributed variables (where the weights can be consistently estimated). The constants* Δ, μ_j *and* $\theta^*_{t_j,T_1}$ *are defined in the Appendix.*

Unfortunately, Proposition 1 cannot be directly applied because the "misspecification term"

$$\sum_{j=1}^{M} n_{2j}\Delta + \sum_{j=1}^{M} \frac{\sqrt{n_{2j}}}{\sqrt{n_{1j}}} \sqrt{n_{2j}}\mu_j' \sqrt{n_1}(\widehat{\theta}_{t_j,T_1} - \theta^*_{t_j,T_1})$$

is not known (nor estimable). Its presence is caused by the transformation of the original nested nonparametric testing problem into a nonnested parametric one. To solve this problem, the unknown constants Δ and μ_j (defined in the Appendix) are assumed to be very small, reflecting the assumption that misspecification of the parametric model for the state-price densities leads to a misspecification error that is small relative to the noise in the data[14]. De Wachter (2003) provides evidence that the "SMN" parametrization in Appendix 6.1 is an empirically sound choice. This provides support for applying Proposition 1 ignoring the misspecification term.

In an extensive series of Monte Carlo experiments, De Wachter (2003) finds that the test for independent returns has actual size close to the nominal one at all convential significance levels. Ignoring the misspecification term does not affect test behaviour if the parametrization is chosen flexible enough. However, overly flexible models may suffer from numerical instability, overly restrictive models from inferior conformance to the distribution theory in Proposition 1.

The power against a realistic calibration of Heston's (1993) stochastic volatility model is close to 100% for datasets spanning 5 days, and remains large even when only a single observation day is used. Power decreases significantly only when pricing errors are calibrated to be 4 times as large as what one typically finds in market data. Details are available upon request.

4 Application to S&P 500 options

The purpose of the test developed above is to examine the performance of simple models over stability periods equal to or slightly greater than the typical recalibration horizon used by traders and risk managers. I therefore organize the discussion of this application by stability period M.

[14]Note, however, that although Δ and μ can reasonably be assumed to be extremely small, the misspecification term is $O\left(\sum_{j=1}^{M} n_{2j}\right)$. Hence, the approximation will, somewhat counterintuitively, only work in small samples.

In this application, the focus is on daily European style S&P 500 option prices for January to June 1993, taken from the well-known "benchmark" dataset in Aït-Sahalia and Lo (1998). The reader is referred to the original paper for a detailed description of the dataset. At the end of the first subsection, more recent but similarly constructed data are used to verify the initial results. These are taken from the IvyDB database[15]. When selecting observation dates t_j and maturity dates T_e, the first maturity selected is at least one month after the last observation date, and the second maturity is roughly one month after the first. All calculations are performed using programs written in the matrix language *Ox* (Doornik (2001)).

4.1 Imposing stability over 1 week

I start by summarizing the main finding. Applying the independence and Markov tests to a single week's options data (i.e. $M = 5$) shows that

1. the test for the Markov property does not lead to a clear-cut rejection, indicating that there *may* exist a very simple option pricing model that can capture the variability in the data without needing recalibration within this week.

2. the test for independence results in a very strong rejection. Interestingly, the main difference between the two is that for the Markov case the estimated forward density $q_{T_1,T_2}(S_{T_2}|S_{T_1})$ shows a strong negative relationship between S_{T_1} and dispersion. This suggests that leverage is an essential ingredient in any model of short-term index options.

The details are as follows. Figure 2 presents the results for a dataset containing all prices of May and June options observed in the week starting on March 22, 1993. This particular week is most suitable for graphical presentation because the available strikes happen to be nicely centred around the money. Each of the 4 plots contains the actual and fitted smiles[16] for June (i.e. T_2) options for a single day. The plot for Friday is similar and is suppressed to avoid clutter. The curve labelled "sep SPD" displays the implied volatilities of option prices estimated using a separate SPD for the cross-section, i.e. estimated under the alternative. The two other curves represent the fitted prices imposing independence and the Markov property, respectively (i.e. estimated under each of the null hypotheses)[17].

The difference in pricing performance between the separate SPDs and those obtained by imposing the Markov property seems extremely small[18]. For independent returns, however, the situation is drastically different. Although fitted implied volatilities for very near-the-money options seem to be on par with those of the other two parametrizations, two striking

[15] See http://www.optionmetrics.com/IvyDB.htm. Cross-sections in this dataset are up to twice as large as those in the Aït-Sahalia and Lo (1998) one.

[16] A "smile" is jargon for the schedule of Black-Scholes implied volatilities of options observed at the same day and maturing at the same date, but with different strike prices. Note that BS implied volatilities are nothing more than a representational device, i.e. they are simply a convenient transformation of prices.

[17] Initial parametrizations include 4 parameters for all SPDs. The constant A (see expression (12)) was set to 0.02.

[18] When interpreting pricing errors from smile-plots, note that sensitivity of Black-Scholes implied volatility error to dollar pricing error is U-shaped across the cross-section.

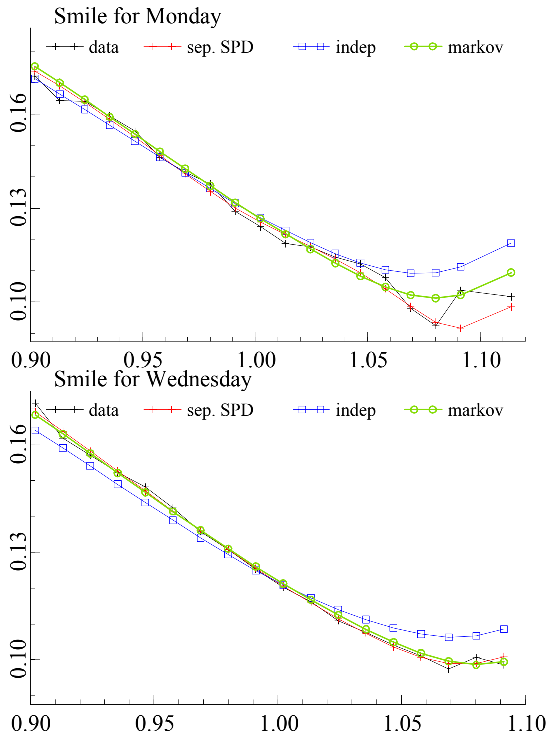

Figure 2: Actual and estimated smiles (Black-Scholes implied volatilities (Y-axis) vs strike / future (X-axis)) for June options observed during the week of March 22, 1993. (May options (not displayed) are used as the first expiry for the independence and Markov cases) The plot for Friday is similar but not shown.

types of mispricing are apparent from the plots. Firstly, ITM calls (or OTM puts) are consistently underpriced. Secondly, OTM calls are consistently overpriced. The overall outcome for the data under study was a convincing rejection of the independent returns property: the value of the test-statistic was 3 times the 99.9^{th} percentage point of the distribution under the null.

Underpricing of ITM and overpricing of OTM calls points to SPDs q_{t_j,T_2} with too little mass in the left tail and too much in the right. This is not an indication that models with independent returns cannot generate conditional densities with strongly negative skewness[19]. The real problem for these models is to fit *both* the term-structure and the cross-section *at the same time*, especially if a certain amount of stability is required. As soon as *any* model with independent returns is forced to price options with a first expiry T_1 correctly, it is unable to allocate enough probability mass in the left tail of the SPD for a later expiry T_2.

This observation points to the crucial shortcoming that is common to all models with independent returns. This shortcoming is most easily understood by comparing the implied SPDs $q_{T_1,T_2}(S_{T_2}|S_{T_1})$ for the independence and the Markov case. Figure 3 plots these SPDs for 3 different values of S_{T_1}. For the case of independent returns, the "shape" of the SPD

[19] Recall that it is always possible (in principle) to find a model with independent increments that can produce a single given SPD (i.e. fits a single cross-section).

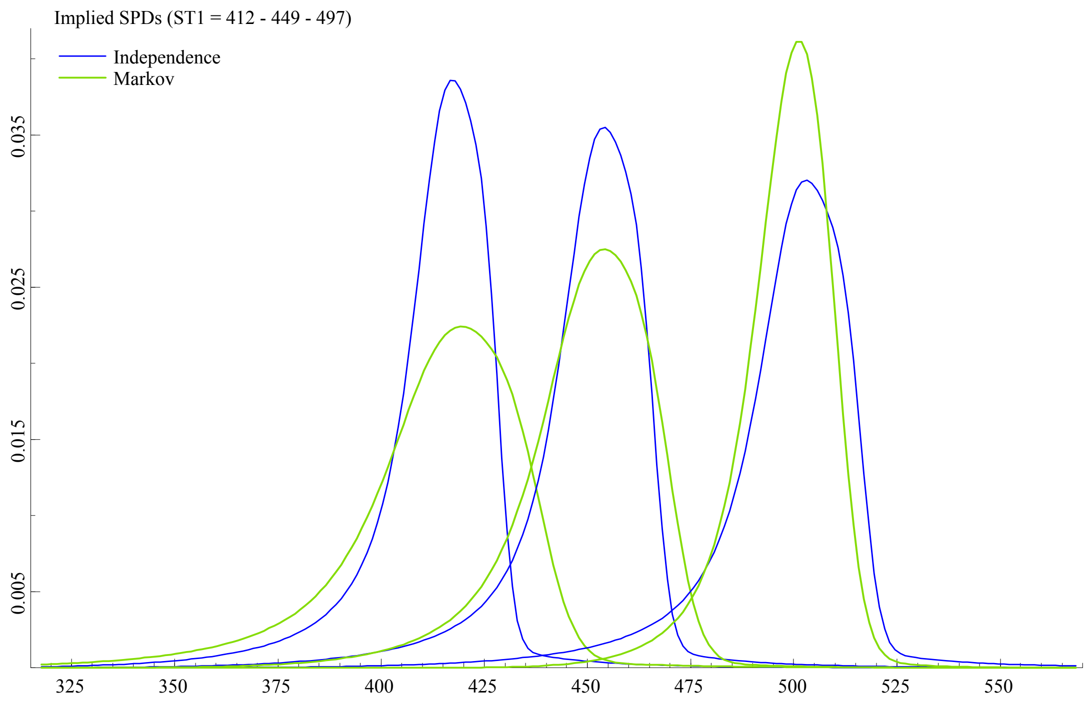

Figure 3: The leverage effect embodied in option prices: estimated implied SPDs q_{T_1,T_2} for 3 different values of S_{T_1} based on May and June options observed during the week of March 22, 1993.

is fixed since the *return* over the period $T_1 - T_2$ is - by definition of independent returns - constrained to have a distribution that does not depend on S_{T_1} (and hence the conditional distribution of the *level* S_{T_2} depends on S_{T_1} in a proportional way). Under the Markov property this constraint is not imposed and this is exactly what allows Markovian models to perform so much better. The Markov-implied SPDs q_{T_1,T_2} have a much greater dispersion for lower values of S_{T_1}. In other words, according to *any* Markov model that will fit the data, volatility over the period $T_1 - T_2$ will be greater the lower the value S_{T_1} of the underlying at the start of that period. Yet otherwise stated, any Markov model that can describe S&P 500 index option prices will have to incorporate a leverage effect. Models with independent returns are by definition unable to produce such a feature and are consequently denied a reasonable fit.

Although the graphical evidence against the Markov property is nonexistent, the statistical results are mixed. Depending on the parameterization used, the p-values range from 0.01 to 0.15. Given the lack of any obvious mispricing pattern, such values do not warrant a rejection.

As a final remark on the results in Figure 2, I point out that the important areas of consistent mispricing by models with independent returns are the wings of the smile (or the tails of the distribution). This suggests that for the purpose of model selection and calibration, it is important to use a large range of strikes. When only prices of options in, say, a 5% region around the money are available, it may become difficult to discriminate between alternatives.

I now repeat the above analysis in the rest of the dataset. Observation dates are selected as follows. Starting with the first Monday in the dataset, I verify that enough data - at least 8 options per cross-section - are available for each day of the week (this is not always the case). Then I select the nearest two expiration dates that are at least a month away and at least a month apart. This procedure is iterated through the entire dataset, producing 16 non-overlapping weeks.

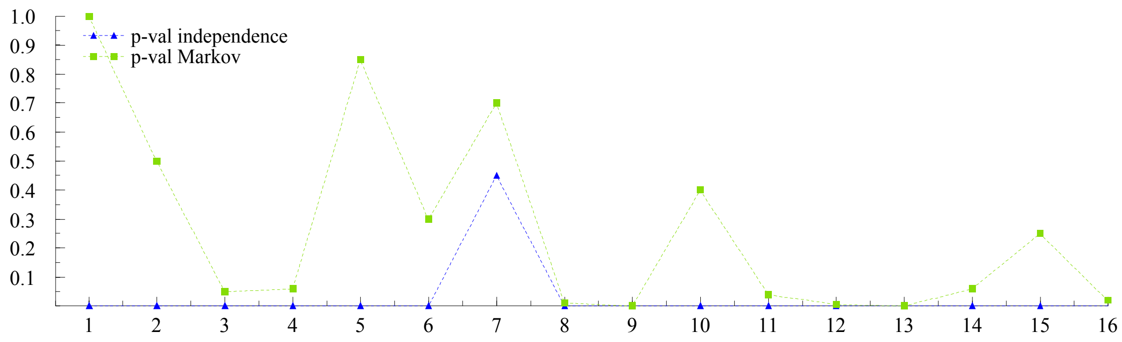

Figure 4: p-values for each qualifying week (S&P 500 options, January 1 to June 30, 1993).

The statistical results for the 5-day case are displayed in Figure 4[20]. The X-axis displays the week-index, not the actual starting date; e.g. week 8 contains the data examined before in section 4.1. The results are roughly in accordance with those obtained before: the independent returns property is conclusively rejected, where evidence against the univariate Markov property is weak. In each case, this can be attributed to the inability of models with independent returns to capture leverage: the forward density plots are qualitatively similar to Figure 3 for each of the observation periods (plots not shown for reasons of space)[21]. The only exception is period 7, where the amount of leverage displayed in the forward densities happens to be somewhat lower.

There is no reason to assume that the observed pattern in the 1993 data - outright rejection of independent returns vs weak evidence against the Markov property - will remain stable over time. As traders' behaviour and quantitative techniques evolve and economic conditions fluctuate, one might encounter changes in the conclusions of my tests. To provide some further evidence, Figure 5 is the counterpart of Figure 4 for data recorded a decade later. The conclusions obtained before appear to remain unchanged. Inspection of individual densities (not shown) revealed the same qualitative picture as for 1993 data.

Overall, one may conclude that no totally conclusive statistical evidence exists either in favour or against univariate Markovian models when a 1-week period of stability is im-

[20]Note that the algorithm used to generate these results automatically reduces the parametrization if it does not manage to find a maximum at a value where the least-squares function is convex. This happens very rarely, but may cause an inferior fit under the alternative, leading to negative values of the test-statistic (see e.g. the first data-block).

[21]The forward density plots under independence often look "jagged": this is additional evidence that the model is unable to fit the data. When univariate Markovianness is imposed this also happens occasionally.

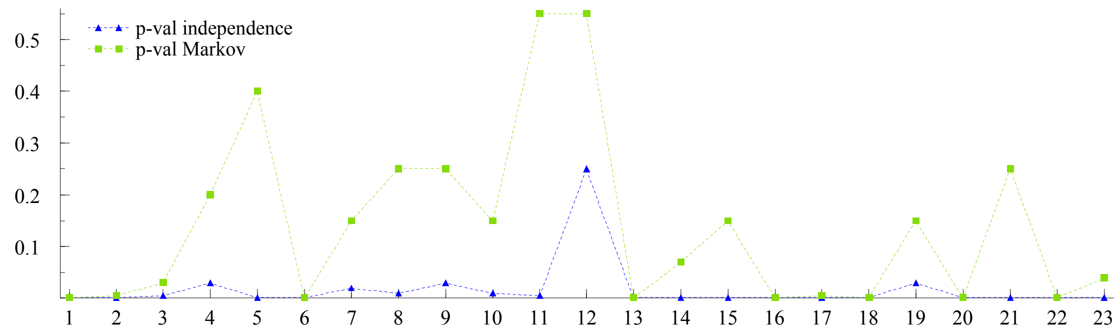

Figure 5: p-values for each qualifying week (S&P 500 options, January 1 to June 30, 2003).

posed. Models with independent returns, on the other hand, have clearly been shown to be inadequate. The next subsection examines the merit of models with independent returns in situations where daily rather than weekly recalibration is allowed.

Before turning to this issue, one may wonder whether stability beyond one week is feasible for univariate Markovian models. As can be expected, the statistical evidence against the independent returns hypothesis grows even stronger as more stability is required. The general pattern of mispricing remains identical to that observed before, but the magnitude of the errors increases.

For the Markov hypothesis, the value of the test-statistic also moves further and further out into the tail as more observation days are added. From a statistical point of view, one cannot maintain the validity of univariate Markov models for observation periods greater than one week (details suppressed). However, even when as many as 20 days are used, no clear returning pattern of mispricing as for the independence case becomes evident.

It is useful to inspect the magnitude of the mispricing in Table 1, which contains the absolute average pricing errors for options in several moneyness categories under 3 different levels of imposed stability. Near the money (NTM) errors are computed by averaging absolute pricing errors for options with strikes in a 5% range around the money. In the money (ITM) errors are based on options with strikes lower than 2.5% below the money, and similarly for OTM errors. Note that these are dollar errors (expressed in cents) and not percentage errors. To give an idea of the magnitudes involved, NTM options are roughly priced around the \$10 mark, ITM call prices range from \$20 to \$50 and far OTM calls can trade for \$3 down to as little as 15 cents.

It is clear from Table 1 that Markov pricing errors increase gradually as more stability is imposed. The statistical test leads us to conclude that pricing errors become too large as soon as more than one week's stability is imposed; the question is whether this is also economically meaningful. Even though I could not detect any consistent patterns of mispricing by eyeballing the smile plots (not shown), the relative increase in pricing errors is fairly substantial.

	one week			two weeks		three weeks	
	sep. SPD	indep	Markov	sep. SPD	Markov	sep. SPD	Markov
ITM	*6.6*	16.6	7.9	*5.7*	7.5	*5.7*	8.5
NTM (5%)	*8.9*	12.9	10.9	*9.4*	12.3	*9.7*	14.5
OTM	*6.5*	22.1	8.3	*6.4*	9.9	*6.1*	10.0
overall	*7.1*	17.6	8.8	*6.8*	9.5	*6.8*	10.4

Table 1: Average absolute pricing errors (in cents) for different moneyness categories and periods of stability; the observation period starts on March 22, 1993

4.2 Independent returns and daily recalibration

Figure 6 presents two instances of the counterpart of Figure 2 for the case of a single observation day[22]. Each of the two panels represent the result of one test (I ran the test twice, i.e. for two different days). First consider the RHS panel (based on data for Tuesday March 23). Clearly, pricing performance suffers significantly when independence is imposed, and the form of the mispricing is similar to that in Figure 2. From inspection of the mispricing pattern, it is apparent that even for the single day case the inability of models with independent returns to capture the leverage effect causes their lack of fit. Hence it can be concluded that even with frequent recalibration, models with independent returns will provide inadequate pricing performance.

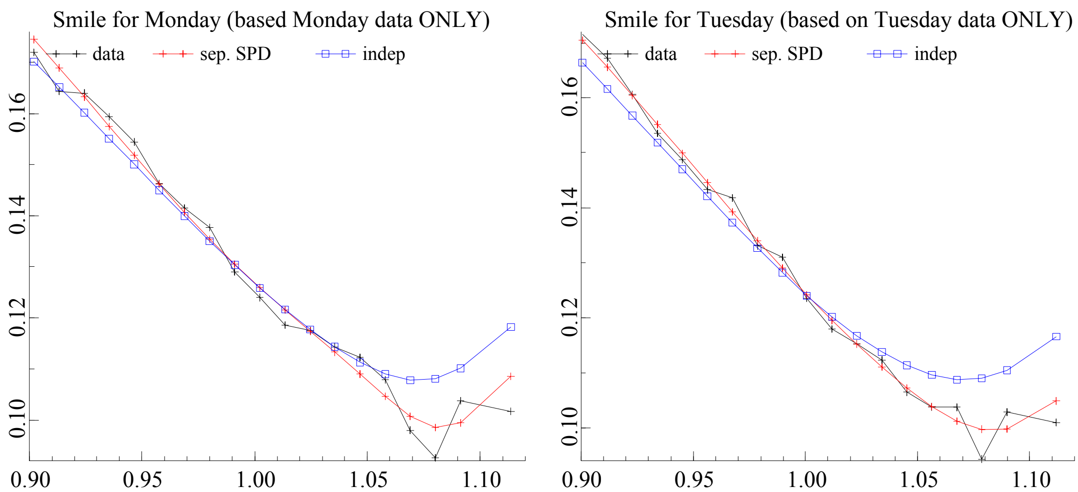

Figure 6: Actual and estimated smiles for June options observed on March 22 and March 23 (unrelated calculations). May options (not displayed) are used for the first expiry in the calculations.

The statistical results confirm the graphical analysis above. The p-value for the independence test was <0.001; however, compared to the 5-day case reported above, the rejection

[22]The fit under the Markov property (not shown) is virtually identical to that obtained using a separate SPD. Indeed, when only a single observation day is used, the Markov test is not interesting for the simple reason that one can always decompose a random variable S_{T_2} into another variable S_{T_1} and the remainder $S_{T_2} - S_{T_1}$ (as long as nothing is imposed on the distribution of the remainder). More generally, it can be proven that there always exists a univariate Markovian model that can fit prices observed at a single point in time.

was far less drastic.

Turning to the LHS panel of Figure 6, the situation seems somewhat less problematic. The p-value in this case was only 0.02, leading to a very marginal rejection at best. With a narrower range of strikes, it would have been very difficult to find any statistical evidence in these data against models with independent returns.

The strength of the evidence against models with independent returns is somewhat surprising, especially given the fairly recent research activity in the area. To rule out the possibility that these results are due to "bad luck", I rerun the test for independent returns for each day in the dataset for which suitable expiration dates could be found (using the same recipe as for the 5-day case). The results are depicted in Figure 7. For most almost every of the 106 available observation days, the independent returns property is rejected.

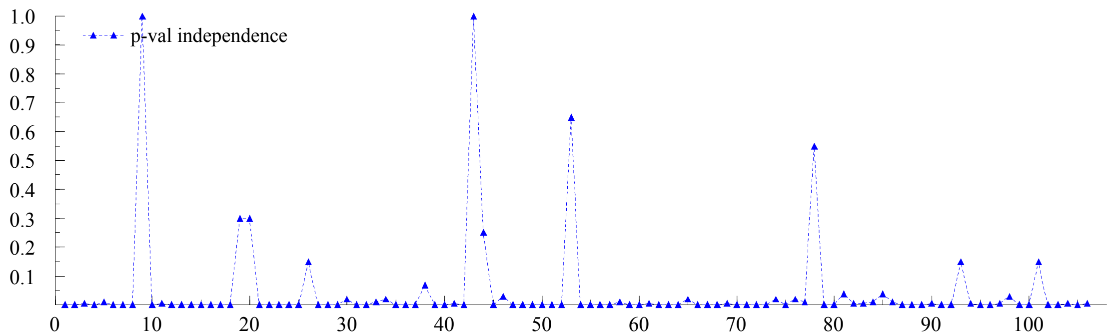

Figure 7: p-values for the independent returns test (S&P 500 options, January 1 to June 30, 1993).

Given these statistical results and the economic interpretation of the causes of mispricing, I can safely conclude that models with independent returns are significantly misspecified and not useful for practical applications, even with frequent recalibration.

It is interesting to contrast this finding with the conclusions of a number of previous studies. Madan, Carr, and Chang (1998), for instance, find that the Variance-Gamma model *"appears to deliver acceptable option prices"* (p.94). This is in sharp contradiction with my results, although admittedly opinions may differ on what is regarded as "acceptable". It may be that the dataset used in Madan, Carr, and Chang (1998) is too small to generate strong evidence against their model[23] or that it contains very noisy prices; it is equally possible that the authors' procedure for analyzing pricing errors is unable to detect the model's inability to capture leverage[24].

[23] Madan, Carr, and Chang (1998) calibrate their model to one week's option data at a time (iterated over each week during a 3-year period), using options with 4 different maturity dates for each day, but only 4 (or less) options per cross-section (i.e. at most 16 options per observation day). If strikes within each cross-section are in a narrow range around the money, it may become more difficult to discriminate between models.

[24] Their conclusion is based on a regression of pricing errors on a constant, moneyness and its square, time-to-maturity, and interest rate. None of the coefficients are statistically significant and the overall explanatory power is very low compared to other specifications. However, consider (for the sake of the argument) what happens when fitting a model with independent returns to two cross-sections of options (i.e. two different maturities) observed on a single day. Since the model cannot capture leverage, it will assign slightly too much

Carr and Wu (2003) develop an option pricing model based on a stable return process and find that, with daily recalibration, it fits the data well and performs on par with a stochastic volatility jump diffusion model in terms of mean-squared pricing error. In particular, they note that this model outperforms Madan, Carr and Chang's Variance-Gamma model because it is designed to capture the extremely slow "flattening out" of the skew with maturity. My results show that, even though Carr and Wu's model may well have superior properties than other models with independent returns, it is not useful for capturing variability in options data. The only way to adequately capture the slow reversion to normality and the associated flat skew as maturity increases is by building in leverage.

5 Conclusion

This paper has developed nonparametric tests for the validity of the univariate Markov property and of independent returns in financial asset-price processes. The method uses only prices of derivative securities as an input and is based on a statistical relationship between transition densities known as the Kolmogorov-Chapman equation. The tests differs from earlier attempts to address the same question in two ways. Firstly, their null hypothesis encompasses a wide variety of models that have appeared in the recent literature. Secondly, they only impose a limited stability period, thereby considering model stability over periods of the same order as those used in practice.

The tests are easy to compute using the parametrization of risk-neutral densities by the smooth mixtures of normals model (or any other convenient method). The drawback of this parametric approach is the appearance of an unknown misspecification term in the asymptotic approximation of the distribution of the test. The advantage of the associated computational simplicity, on the other hand, is the possibility to examine the behaviour of the tests in simulation experiments (for the independent returns test). These experiments (not included in this paper) revealed regular behaviour under the null and excellent power properties under realistic conditions.

The results of applying the tests to S&P 500 data go against the often held belief that univariate Markovian models are not "sophisticated" enough for pricing purposes. As long as one is willing to recalibrate models on a weekly basis, evidence in the forward properties of the data against this class of processes is unconvincing at best. In contrast, models with independent returns do not perform satisfactorily. Although it is certainly the case that they can generate enough skewness to "capture the smile", i.e. fit a single cross-section of option prices observed at a single point in time, they are unable to simultaneously fit the term-structure (several cross-sections with different maturities). The reason for this failure is their inability to incorporate leverage. This finding is not in agreement with the conclusions

mass to the left tail for the first maturity such that the second maturity suffers less of a deficit in that area. The reverse will happen for the right tail. As a result, ITM calls (or OTM puts) for the first maturity will be slightly overpriced and second maturity ITM calls slightly underpriced (vice-versa for OTM calls). Because no interaction terms between moneyness and time-to-maturity are included, the regression analysis of Madan, Carr, and Chang (1998) will not reveal a relationship between moneyness and pricing errors because there will be roughly as many overpriced as underpriced options for each moneyness level (pricing errors for all maturities are pooled).

of a number of other studies. Additionally, it stresses the role of volatility asymmetry in generating skewness in asset returns.

The reader may wonder whether the focus of my tests on short periods of model stability is relevant from an academic point of view. After all, recalibration is not logically consistent with the theory underlying the pricing models! It certainly would not be relevant if methods were available that achieve comparable levels of pricing accuracy as those reported here whilst maintaining stability over several months or years. As I have shown, univariate Markovian models are definitely not able to do so. However, examinations of much more general specifications such as the bivariate stochastic volatility model developed in Bates (2000) also indicate considerable misspecification. Until fully stable models are discovered, then, the question is what the desired stability period is to be. Certainly, opinions on this issue will differ among academics and practicioners, especially when computational speed is important. What I have shown here is that univariate Markovian models should not be expected to perform over periods of more than a week.

Even if the inability of any model examined to date to satisfactorily price all options observed over a long period of time may prompt the development of even more flexible specifications, it is perhaps just too ambitious to aim for a single "correct" model. Especially when computational speed is an issue, the relevant question may instead be to select the least misspecified model for the given purposes. For instance, in order to price long maturity options, it may be useful to introduce stochastic interest rates and stochastic volatility into the model (e.g. Bakshi, Cao, and Chen (2000)). For the pricing of short-term options, such added flexibility may well not offset the increase in computational complexity since other much simpler models can fit the cross-section and term structure of option prices in this case as well. In practical situations, a slight lack of stability of such models can be dealt with by recalibration.

6 Appendix

6.1 Parametric specifications

Parametric estimation of state-price densities is based on the so-called "*Smooth Mixture of Normals*" (SMN) model, which imposes some smoothness whilst maintaining the flexibility of mixture models. The idea is to formulate a uniform mixture of normals

$$q_{t,T}(S_T|I_t) = \int_0^1 LN(\delta(\alpha), s(\alpha))d\alpha$$

where $LN(\delta(\alpha), s(\alpha))$ is shorthand notation for the lognormal density with mean and volatility parameters $\delta(\alpha)$ and $s(\alpha)$. The latter are constructed as smooth functions of α, for reasons discussed below.

For computational simplicity I implement the idea as a discrete mixture by parametrizing

$q_{t,T}(S_T|I_t)$ as a smooth mixture of $l+1$ lognormals[25]

$$q_{t,T}(S_T|I_t) = \frac{1}{l+1}\sum_{i=0}^{l}\frac{1}{S_T\sqrt{2\pi s_i^2}}\exp\left\{-\frac{\left[\log(\frac{S_T}{S_t}) - (r-d)(T-t) - (\delta_i - \frac{s_i^2}{2})\right]^2}{2s_i^2}\right\} \quad (11)$$

with

$$\begin{aligned} \delta_i &= (T-t).\left(b_0 + b_1(\frac{i}{l} - 1/2)\right) \\ s_i &= \sqrt{T-t}\ \left(A + \exp\left\{\sigma_0 + \sum_{j=1}^{p}\theta_j.(\frac{i}{l} - 1/2)^j\right\}\right) \end{aligned} \quad (12)$$

where $(b_1, \sigma_0, \theta_1, ..., \theta_p)' = \theta$ are parameters to be estimated and b_0 is chosen so as to satisfy the martingale constraint, which in this case means

$$b_0 = -\frac{1}{T-t}\log\left[\frac{1}{l+1}\sum_{i=0}^{l}\exp\left(b_1(\frac{i}{l} - 1/2)(T-t)\right)\right]$$

Note that the specification of s_i contains a positive constant A that rules out excessively sharp "spikes" in the SPD. This functional form can be interpreted as the transition density of a mixture over $l+1$ geometric Brownian motions with drifts $\delta_i/(T-t)$ and volatilities $s_i/\sqrt{T-t}$.

Bahra (1997), Melick and Thomas (1997) and Ritchey (1990) present similar parametrizations (i.e. based on mixtures of (log-)normals). The distinctive feature of the SMN approach is that the different subdensities are linked through the smooth functions δ_i and s_i. This means that subdensities i and j with adjacent means cannot differ by too much (since s_i and s_j cannot differ very much), ruling out strong "kinks". Bahra (1997) and Jondeau and Rockinger (2000) report potential problems with "spikes" when testing another method based on mixtures of lognormals. Note that one could give this approach a "seminonparametric" flavour by letting the degree p of the polynomial in s_i vary as the number of datapoints changes. De Wachter (2003) derives asymptotic rates of convergence for a related mixture model; proving consistency for the SMN model itself has, however, not been possible.

A major advantage of the SMN parametrization in computational terms is that the associated call-pricing formula is available in closed form as the following sum of Black-Scholes-like expressions

$$C_{t,T}(K;\theta) = \frac{1}{l+1}\sum_{i=0}^{l}\left\{S_t\ e^{(\delta_i - d(T-t))}\ \Phi(\gamma_i + s_i) - e^{-r(T-t)}\ K\ \Phi(\gamma_i)\right\} \quad (13)$$

[25]Depending on the computation speed required, n can be set appropriately. In my experience this does not have a noticeable effect.

where $\Phi(.)$ is the standard normal distribution function and

$$\gamma_i = \frac{\log(\frac{S_t}{K}) + (r-d)(T-t) + (\delta_i - \frac{s_i^2}{2})}{s_i}$$

The parameters can be estimated by nonlinear least squares on the model

$$C_{t,T}^{K_i} = C_{t,T}(K;\theta) + \varepsilon_i \quad i = 1, ..., n$$

where $C_{t,T}(K, \theta)$ is given by (13). Distribution theory for misspecified NLS to deal with the measurement error ε is developed in White (1981).

The regression function $C_{t_j,T_2}^{Ind}(K_{t_j,T_2}; \widehat{\theta}_{t_j,T_1}, \theta_{Ind})$ in (9) contains an integral (see (2)). An attractive feature of the SMN parametrization is that calculation of this integral is easy (for the case of independent returns). To see why, let $\phi_X(.)$ denote the characteristic function of the random variable X and note that if X and Y are independent and $Z = X + Y$, the convolution theorem states that $\phi_Z(m) = \phi_X(m).\phi_Y(m)$.

To simplify notation, denote the time-t price of a future expiring at T by $F_{t,T}$. Reformulating our parametrization (11) of the SPD in terms of log-prices[26] and indicating with a superscript-X quantities related to $\log(S_{T_1}) - \log(S_t)$ (e.g. $d_j^X = \delta_j^X - \frac{s_j^{2X}}{2}$) and with superscript-$Y$ those related to $\log(S_{T_2}) - \log(S_{T_1})$, we can write ($i^2 = -1$)

$$\begin{aligned} \phi_X(m) &= \int_{-\infty}^{\infty} e^{imx} \frac{1}{l+1} \sum_{j=0}^{l} \frac{1}{\sqrt{2\pi s_j^{2X}}} \exp\left\{ -\frac{\left[x - d_j^X\right]^2}{2 s_j^{2X}} \right\} dx \\ &= \frac{1}{l+1} \sum_{j=0}^{l} \left[\int_{-\infty}^{\infty} e^{imx} \frac{1}{\sqrt{2\pi s_j^{2X}}} \exp\left\{ -\frac{\left[x - d_j^X\right]^2}{2 s_j^{2X}} \right\} dx \right] \\ &= \frac{1}{l+1} \sum_{j=0}^{l} \exp\left(i d_j^X m - \frac{1}{2} s_j^{2X} m^2 \right) \end{aligned}$$

In the first step we defined $x \equiv \log(\frac{S_{T_1}}{F_{t,T_1}})$ as the demeaned continuously compounded return[27]. The third step is a known identity. Similarly

$$\phi_Y(m) = \frac{1}{l+1} \sum_{k=0}^{l} \exp\left(i d_k^Y m - \frac{1}{2} s_k^{2Y} m^2 \right)$$

[26] I.e. in the above notation, replace X by $\log(S_{T_1}) - \log(S_t)$, Y by $\log(S_{T_2}) - \log(S_{T_1})$ and Z by $\log(S_{T_2}) - \log(S_t)$.

[27] $\log(\frac{S_{T_1}}{F_{t,T_1}}) = \log(S_{T_1}) - \log(S_t) - (r-d)(T_1 - t)$

Combining these gives

$$
\begin{aligned}
\phi_Z(m) &= \left[\frac{1}{l+1}\sum_{j=0}^{l}\exp\left(id_j^X m - \frac{1}{2}s_j^{2X}m^2\right)\right]\left[\frac{1}{l+1}\sum_{k=0}^{l}\exp\left(id_k^Y m - \frac{1}{2}s_k^{2Y}m^2\right)\right] \\
&= \left(\frac{1}{l+1}\right)^2\sum_{j=0}^{l}\sum_{k=0}^{l}\exp\left(i(d_j^X + d_k^Y)m - \frac{1}{2}(s_j^{2X} + s_k^{2Y})m^2\right)
\end{aligned}
$$

It may be helpful to note that

$$
\begin{aligned}
x + y &= \log(S_{T_2}) - \log(S_{T_1}) - (r-d)(T_2 - T_1) + \log(S_{T_1}) - \log(S_t) - (r-d)(T_1 - t) \\
&= \log(S_{T_2}) - \log(S_t) - (r-d)(T_2 - t) = \log(\frac{S_{T_2}}{F_{t_j,T_2}})
\end{aligned}
$$

$\phi_Z(m)$ results in the following expression for the "independence-implied" density of S_{T_2} given S_{t_j}

$$
q_{t_j,T_2}(S_{T_2}|S_{t_j}) = \left(\frac{1}{l+1}\right)^2\sum_{j=0}^{l}\sum_{k=0}^{l}\frac{1}{S_{T_2}\sqrt{2\pi(s_j^{2X}+s_k^{2Y})}}\exp\left\{-\frac{\left[\log(\frac{S_{T_2}}{F_{t_j,T_2}}) - (d_j^X + d_k^Y)\right]^2}{2(s_j^{2X}+s_k^{2Y})}\right\} \qquad (14)
$$

The associated call-pricing function is now a mixture of Black-Scholes-like functions, similar to that in expression (13). This simplification makes the test for independent returns much easier to implement than the Markov test, for which the integrals in system (5) will have to be calculated numerically. Note that the density function $q_{t_j,T_2}(S_{T_2}|S_{t_j})$ in (14) is not an SMN density: this illustrates how the initial nested testing problem is transformed in a non-nested one.

In the Markov case, the dependence of $q_{T_1,T_2}(S_{T_2}|S_{T_1})$ on S_{T_1} can be made explicit by letting θ vary as a smooth function of S_{T_1}, here implemented as a (component-by-component) polynomial. As a practical matter $\theta(S_{T_1})$ will be kept constant for $q_{t_j,T_1}(S_{T_1}|S_{t_j})$-improbable values of S_{T_1}, since for these values $q_{T_1,T_2}(S_{T_2}|S_{T_1})$ does not affect the option price enough to be numerically relevant. The dependence is represented schematically in Figure 8.

6.2 Derivation of asymptotic distribution theory

In this appendix, I explain the derivation of Proposition 1. Because of the complicated structure of the data, one of the main difficulties is notation. For the purpose of clarity, I consider a simplified situation with only a single observation day. I also simplify the notation. Full details of the general case are available from the author upon request.

6.2.1 Setup and notation

The sample consists of $n_1 + n_2$ observations, grouped in the vectors y_1 and y_2, corresponding to T_1 and T_2 option prices (for a single observation date) respectively. Let $w(\lambda)$ and $z(\gamma)$ denote (n_1 and n_2-vectors of) functions (call-pricing functions each derived from a single

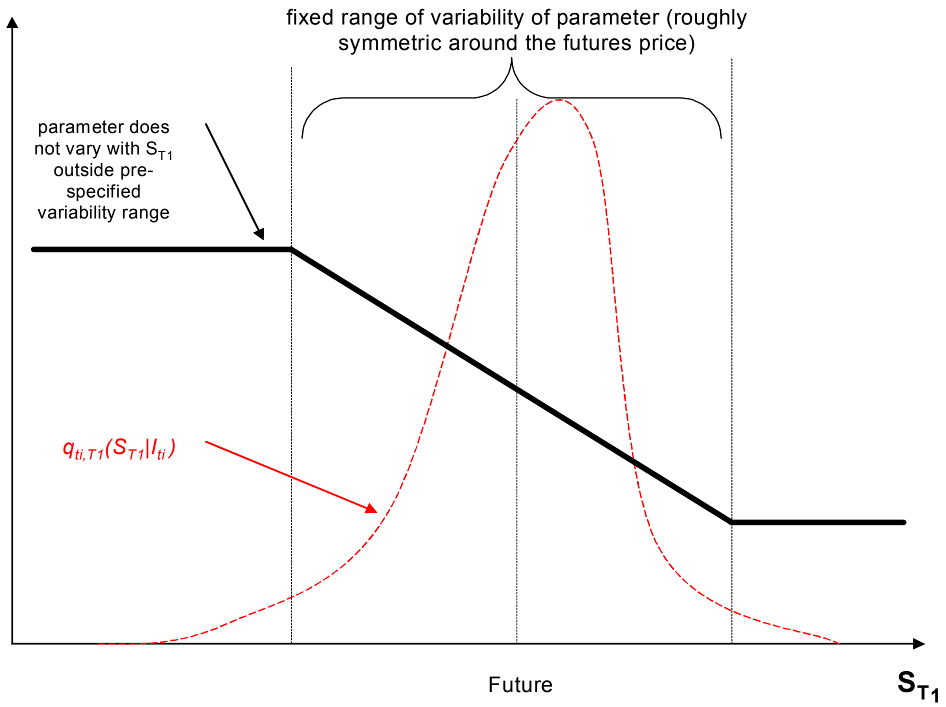

Figure 8: The functional form of the dependence of the parameters of the implied Markov call pricing function $C^{K_j}_{T_1,T_2}(S_{T_1})$ (and SPD $q_{T_1,T_2}(S_{T_2}|S_{T_1})$) on the intermediate value of the underlying. The thick line is represents a parameter value. The slope and intercept of the middle part are estimated; outside this interval the parameter is kept constant at the value attained at the endpoints of the interval. The variability range is fixed (typically at 20% of the futures value) and symmetric around the futures value of the middle observation day in the sample.

SMN SPD) corresponding to T_1 and T_2 options as in expression (7). $x(\beta, \lambda)$ corresponds to the pricing function for T_2-calls constructed using the Kolmogorov-Chapman equation (as in (8)). It contains the same covariates as $z(.)$. These covariates K (the strikes) are considered to be fixed or exogenous and are suppressed in the notation. The model pricing errors are represented by the n_1, n_2 and n_2-vectors v, η and ζ respectively. Observations are indexed by the (second in some cases) subscript-i. Parameter vectors $\lambda \in R^{l\times 1}$, $\beta \in R^{b\times 1}$ and $\gamma \in R^{g\times 1}$ are not known.

I want to test the null hypothesis

$$H_0 : \begin{array}{c} y_1 = w(\lambda) + v \\ y_2 = x(\beta, \lambda) + \eta \end{array} \tag{15}$$

against the alternative

$$H_1 : \begin{array}{c} y_1 = w(\lambda) + v \\ y_2 = z(\gamma) + \zeta \end{array} \tag{16}$$

Note that only the second equation differs under both hypotheses.

The call-pricing functions in both H_0 and H_1 are likely to be misspecified. Therefore, it is not necessarily the case that any of the models H_0 or H_1 contain the DGP as a special case. Using the same vector notation as above, this DGP is given by

$$DGP: \begin{array}{l} y_1 = w^0 + \varepsilon_1 \\ y_2 = z^0 + \varepsilon_2 \end{array} \tag{17}$$

Here ε_1 and ε_2 are (component-by-component) independent pricing errors with the property that $E(\varepsilon_{ei}) = 0$ and $E(\varepsilon_{ei}^2) = \sigma^2 < \infty$; $e = 1, 2$, $i = 1, ..., n_e$.

The standard machinery for discriminating between H_0 and H_1 in this parametric set-up is that of testing for encompassing (see Gourieroux and Monfort (1995)). Intuitively, H_0 encompasses H_1 if the former can "explain all the features of the data that H_1 can". Encompassing tests are designed to deal with situations where very different (non-nested) explanations of a phenomenon are competing for the status of "best theory". The present problem, however, is different in spirit. The question I ask is whether the true call pricing function for T_2 options, z^0, can be "decomposed" into two parts, one of which is w^0, the other related to the SPD $q_{T_1,T_2}(S_{T_2}|S_{T_1})$. This is a restriction on the "most general model" and thus a nested testing problem. Indeed, the original alternative hypothesis can be described as "option prices are rationalized by *some* martingale measure" and the original null hypothesis reads "this martingale measure satisfies independent returns (the univariate Markov property)". Clearly, these statements are nested, but do not imply a parametric functional form for the call-pricing functions.

The problem, then, arises because on the one hand the functional form of z^0 is unknown, but on the other the small sample size of about 10 to 25 options per cross-section rules out a fully nonparametric estimation procedure. By parametrizing the call-pricing functions I introduce both misspecification and nonnestedness, transforming the originally nested nonparametric problem into a nonnested parametric one. As explained in detail below, I proceed by ignoring most aspects of the non-nestedness of the resulting parametric problem.

The success of the approach introduced below hinges crucially on the assumption that the "amount of misspecification and nonnestedness" be small under the null. In other words, I assume that $w(\lambda)$ and $z(\gamma)$ (and under the null also $x(\beta, \lambda)$) approximate w^0 and z^0 accurately. Not all aspects of this assumption are testable. A specification test for separately estimated call-pricing functions based on general-purpose nonparametric specification tests in Zheng (1996) is presented in De Wachter (2003). Below, I derive expressions capturing the non-testable aspects of misspecification and nonnestedness; in an extensive series of experiments with artificial data (De Wachter (2003)) these quantities turn out to be negligible when using a sufficiently flexible SMN model.

6.2.2 Pseudo-true values

According to standard theory (White (1981)), the probability limits of the NLS estimators of λ, β, and γ are the pseudo-true values λ^*, β^*, and γ^* defined as

$$\lambda^* = \arg\min_{\lambda} \int (y_1 - w(\lambda))^2 dF_0 \tag{18}$$

$$\beta^* = \arg\min_{\beta} \int (y_2 - x(\beta, \lambda^*))^2 dF_0 \tag{19}$$

$$\gamma^* = \arg\min_{\gamma} \int (y_2 - z(\gamma))^2 dF_0 \tag{20}$$

F_0 is the distribution of y and the covariates corresponding to the DGP.

Note that the definition of β^*, the pseudo-true value of the intermediate SPD q_{T_1,T_2} depends on whether it is estimated jointly with λ. The fact that I estimate the parameters in model H_0 by first estimating λ from the first equation in (15) and then plug in the resulting estimate in the second equation to estimate β is reflected in (19).

Remark 1 *(Nonparametric vs parametric) Using the notation just introduced, one can be more precise about the nature of the problems introduced by the parametric approach. Firstly,*

$$\textit{it is not necessarily true that} \quad \lambda^* = \arg\min_{\lambda} \int (y_2 - x(\beta^*, \lambda))^2 dF_0 \tag{21}$$

This equation ***does*** *hold if there is no misspecification and the process truly has independent returns (e.g. suppose the SPDs are correctly specified using some stable density function). However, if the process has independent returns but the SPDs are misspecified (as with the SMN parametrization), then the KC equation will not necessarily hold* ***even at the pseudo-true values****.*
Secondly, in general both $x(\beta^*, \lambda^*) \neq z^0$ and $z(\gamma^*) \neq z^0$. These differences are assumed to be extremely small.

6.2.3 Asymptotic distribution

Based on the framework and ideas outlined above, the actual derivation of the distribution test-statistic (10) is relatively straightforward. Some background on misspecified nonlinear regression models can be found in White (1981). Vuong (1989) develops a battery of nested and non-nested tests for parametric models estimated by Maximum Likelihood. The present derivation uses ideas from both papers. Note that no dependence is present between random quantities and therefore only the simplest Central Limit Theorems will be needed.

I first find the first-order asymptotic expansions of the least-squares functions corresponding to each equation in (15) and (16). Using these, I write the difference between squared errors from the second equation in each system (i.e. the test-statistic) as a sum of quadratic forms in normal variables (up to first order) and a misspecification term. The distribution of the former can be estimated; the latter will be assumed to be negligible for practical purposes.

Define

$$ssr^{\lambda}(\lambda) \equiv \sum_i (y_{1i} - w_i(\lambda))^2$$

and similarly $ssr^{\beta}(\beta, \lambda) \equiv \sum_i (y_{2i} - x_i(\beta, \lambda))^2$ and $ssr^{\gamma}(\gamma) \equiv \sum_i (y_{2i} - z_i(\gamma))^2$. The (LR-type) test-statistic (10) is in this notation written as $ssr^{\beta}(\widehat{\beta}) - ssr^{\lambda}(\widehat{\gamma})$. A single term in a sum of this type is referred to by using subscript-i, e.g. $ssr_i^{\lambda}(\lambda) \equiv (y_{1i} - w_i(\lambda))^2$.

I start by examining the sum-of-squares function corresponding to T_2 options priced by a separate SPD. First deal with $ssr^{\gamma}(\widehat{\gamma})$ by performing an expansion around $\widehat{\gamma}$:

$$\begin{aligned} ssr^{\gamma}(\gamma^*) &= ssr^{\gamma}(\widehat{\gamma}) + \left[\frac{\partial}{\partial \gamma'} ssr^{\gamma}(\widehat{\gamma})\right].(\gamma^* - \widehat{\gamma}) \\ &\quad + \frac{1}{2}\sqrt{n_2}(\widehat{\gamma} - \gamma^*)' \left[\frac{\partial^2}{\partial \gamma \partial \gamma'} \frac{1}{n_2} ssr^{\gamma}(\widehat{\gamma})\right] \sqrt{n_2}(\widehat{\gamma} - \gamma^*) + o_p(1) \end{aligned}$$

The second term on the RHS is zero by definition: it is (($\gamma^* - \widehat{\gamma}$) times) the gradient of the LS function evaluated at the estimated value. Similarly for $ssr^{\beta}(\widehat{\beta})$:

$$\begin{aligned} ssr^{\beta}(\beta^*, \lambda^*) &= ssr^{\beta}(\widehat{\beta}, \widehat{\lambda}) + \left[\frac{\partial}{\partial \beta'} ssr^{\beta}(\widehat{\beta}, \widehat{\lambda})\right].(\beta^* - \widehat{\beta}) \\ &\quad + \frac{\sqrt{n_2}}{\sqrt{n_1}}\sqrt{n_2}\left[\frac{\partial}{\partial \lambda'} \frac{1}{n_2} ssr^{\beta}(\widehat{\beta}, \widehat{\lambda})\right].\sqrt{n_1}(\lambda^* - \widehat{\lambda}) \end{aligned}$$

$$\begin{aligned} &+ \frac{1}{2}\sqrt{n_2}(\widehat{\beta} - \beta^*)' \left[\frac{\partial^2}{\partial \beta \partial \beta'} \frac{1}{n_2} ssr^{\beta}(\widehat{\beta}, \widehat{\lambda})\right] \sqrt{n_2}(\widehat{\beta} - \beta^*) \\ &+ \frac{\sqrt{n_2}}{\sqrt{n_1}}\sqrt{n_1}(\widehat{\lambda} - \lambda^*)' \left[\frac{\partial^2}{\partial \lambda \partial \beta'} \frac{1}{n_2} ssr^{\beta}(\widehat{\beta}, \widehat{\lambda})\right] \sqrt{n_2}(\widehat{\beta} - \beta^*) \end{aligned}$$

$$+ \frac{1}{2}\frac{n_2}{n_1}\sqrt{n_1}(\widehat{\lambda} - \lambda^*)' \left[\frac{\partial^2}{\partial \lambda \partial \lambda'} \frac{1}{n_2} ssr^{\beta}(\widehat{\beta}, \widehat{\lambda})\right] \sqrt{n_1}(\widehat{\lambda} - \lambda^*) + o_p(1) \tag{22}$$

The second term on the RHS is zero as above, but this is not necessarily the case for the third term (see (21)). A Taylor expansion around $\widehat{\beta}$ gives after rearranging

$$\begin{aligned} \sqrt{n_2}\frac{\partial}{\partial \lambda}\frac{1}{n_2} ssr^{\beta}(\widehat{\beta}, \widehat{\lambda}) &= \sqrt{n_2}\frac{\partial}{\partial \lambda}\frac{1}{n_2} ssr^{\beta}(\beta^*, \lambda^*) + \left[\frac{\partial^2}{\partial \lambda \partial \beta'} \frac{1}{n_2} ssr^{\beta}(\widehat{\beta}, \widehat{\lambda})\right] \sqrt{n_2}(\widehat{\beta} - \beta^*) \\ &\quad + \frac{\sqrt{n_2}}{\sqrt{n_1}}\left[\frac{\partial^2}{\partial \lambda \partial \lambda'} \frac{1}{n_2} ssr^{\beta}(\widehat{\beta}, \widehat{\lambda})\right] \sqrt{n_1}(\widehat{\lambda} - \lambda^*) + o_p(1) \end{aligned} \tag{23}$$

I now continue the discussion of the effects of the parametric approach.

Remark 2 *(Nonparametric vs parametric) The first of the two problems mentioned in Remark 1 materializes in expression (23). The components of the sum* $\frac{\partial}{\partial \lambda}\frac{1}{n_2} ssr^{\beta}(\beta^*, \lambda^*)$ *do not necessarily have mean zero (they only do if expression (21) holds). Consequently,* $\frac{\partial}{\partial \lambda}\frac{1}{n_2} ssr^{\beta}(\beta^*, \lambda^*)$ *is* $O_p(1)$ *instead of* $O_p(n_2^{-1/2})$ *and it is not possible to apply a CLT directly.*
I proceed by calling its mean (per observation) μ:

$$\mu \equiv E_{F_0}\left(\frac{\partial}{\partial \lambda} ssr_i^{\beta}(\beta^*, \lambda^*)\right)$$

Subtracting this, one can apply a CLT. Unfortunately, μ *is not known and I will make the assumption it is small enough to be negligible in practical applications if the underlying process has independent returns.*
The second problem referred to in Remark 1 is that the expansion for the test-statistic

$ssr^\beta(\widehat{\beta},\widehat{\lambda}) - ssr^\gamma(\widehat{\gamma})$ contains the term $ssr^\beta(\beta^*,\lambda^*) - ssr^\gamma(\gamma^*)$: both terms in this sum are $O_p(n_2)$ but do not cancel each other out as would be the case under correct specification. In principle this makes all the other terms in the expression negligible and one would have to scale down the test-statistic to obtain a well-defined asymptotic distribution (e.g. see the case of strictly non-nested hypotheses in Vuong (1989)). In my setup T_2-SPDs estimated under the null and the alternative should be virtually identical if they are specified flexibly enough. Consequently, I assume that $ssr^\beta(\beta^*,\lambda^*) - ssr^\gamma(\gamma^*)$ is small: call it Δ

$$\Delta = plim \frac{1}{n_2}\left(ssr^\beta(\beta^*,\lambda^*) - ssr^\gamma(\gamma^*)\right) \tag{24}$$

Summarizing, the problem with the parametric approach is that one cannot validly make some asymptotic argument in which the above problems vanish. Instead I rely on the assumption that the ability of the SMN model to capture all practically relevant SPDs ensures that both μ and Δ are very small relative to the sampling error component of the test-statistic.

I now proceed by collecting terms (suppressing the argument of all $ssr-$functions): first use (23) to rewrite (22):

$$ssr^{\beta*} = \widehat{ssr}^\beta - \frac{\sqrt{n_2}}{\sqrt{n_1}}\left\{\begin{array}{c}\sqrt{n_2}\frac{\partial}{\partial\lambda}\frac{1}{n_2}ssr^{\beta*} + \left[\frac{\partial^2}{\partial\lambda\partial\beta'}\frac{1}{n_2}\widehat{ssr}^\beta\right]\sqrt{n_2}(\widehat{\beta}-\beta^*) \\ +\frac{\sqrt{n_2}}{\sqrt{n_1}}\left[\frac{\partial^2}{\partial\lambda\partial\lambda'}\frac{1}{n_2}\widehat{ssr}^\beta\right]\sqrt{n_1}(\widehat{\lambda}-\lambda^*) + o_p(1)\end{array}\right\}' . \sqrt{n_1}(\widehat{\lambda}-\lambda^*)$$

$$+\frac{1}{2}\sqrt{n_2}(\widehat{\beta}-\beta^*)'\left[\frac{\partial^2}{\partial\beta\partial\beta'}\frac{1}{n_2}\widehat{ssr}^\beta\right]\sqrt{n_2}(\widehat{\beta}-\beta^*)$$
$$+\frac{\sqrt{n_2}}{\sqrt{n_1}}\sqrt{n_1}(\widehat{\lambda}-\lambda^*)'\left[\frac{\partial^2}{\partial\lambda\partial\beta'}\frac{1}{n_2}\widehat{ssr}^\beta\right]\sqrt{n_2}(\widehat{\beta}-\beta^*)$$

$$+\frac{1}{2}\frac{n_2}{n_1}\sqrt{n_1}(\widehat{\lambda}-\lambda^*)'\left[\frac{\partial^2}{\partial\lambda\partial\lambda'}\frac{1}{n_2}\widehat{ssr}^\beta\right]\sqrt{n_1}(\widehat{\lambda}-\lambda^*) + o_p(1)$$

Simplifying further:

$$ssr^{\beta*} = \widehat{ssr}^\beta - \frac{\sqrt{n_2}}{\sqrt{n_1}}\left(\sqrt{n_2}\left(\frac{1}{n_2}\sum_i(\frac{\partial}{\partial\lambda}ssr_i^{\beta*} - \mu)\right)\right)'\sqrt{n_1}(\widehat{\lambda}-\lambda^*) - \frac{\sqrt{n_2}}{\sqrt{n_1}}\sqrt{n_2}\mu'\sqrt{n_1}(\widehat{\lambda}-\lambda^*)$$

$$+\frac{1}{2}\sqrt{n_2}(\widehat{\beta}-\beta^*)'\left[\frac{\partial^2}{\partial\beta\partial\beta'}\frac{1}{n_2}\widehat{ssr}^\beta\right]\sqrt{n_2}(\widehat{\beta}-\beta^*)$$
$$-\frac{1}{2}\frac{n_2}{n_1}\sqrt{n_1}(\widehat{\lambda}-\lambda^*)'\left[\frac{\partial^2}{\partial\lambda\partial\lambda'}\frac{1}{n_2}\widehat{ssr}^\beta\right]\sqrt{n_1}(\widehat{\lambda}-\lambda^*) + o_p(1)$$

Note that the cross-product term has disappeared. This is because λ and β are not estimated jointly.

I now collect all terms. Note that all matrices (square brackets) are $O_p(1)$. Further, the difference between e.g. $\left[\frac{\partial^2}{\partial\beta\partial\beta'}\frac{1}{n_2}\widehat{ssr}^\beta\right]$ and $\left[\frac{\partial^2}{\partial\beta\partial\beta'}\frac{1}{n_2}ssr^{\beta*}\right]$ is $o_p(1)$; hence replacing the appropriate quantities we get:

$$\left(\widehat{ssr}^{\beta}-\widehat{ssr}^{\gamma}\right)-\left\{n_2\Delta+\frac{\sqrt{n_2}}{\sqrt{n_1}}\sqrt{n_2}\mu'\sqrt{n_1}(\widehat{\lambda}-\lambda^*)\right\}$$

$$=\frac{\sqrt{n_2}}{\sqrt{n_1}}\left(\sqrt{n_2}\left(\frac{1}{n_2}\sum_i(\frac{\partial}{\partial\lambda}ssr_i^{\beta*}-\mu)\right)\right)'\sqrt{n_1}(\widehat{\lambda}-\lambda^*)$$

$$-\frac{1}{2}\sqrt{n_2}(\widehat{\beta}-\beta^*)'\left[\frac{\partial^2}{\partial\beta\partial\beta'}\frac{1}{n_2}ssr^{\beta*}\right]\sqrt{n_2}(\widehat{\beta}-\beta^*)+\frac{1}{2}\sqrt{n_2}(\widehat{\gamma}-\gamma^*)'\left[\frac{\partial^2}{\partial\gamma\partial\gamma'}\frac{1}{n_2}ssr^{\lambda*}\right]\sqrt{n_2}(\widehat{\gamma}-\gamma^*)$$

$$+\frac{1}{2}\frac{n_2}{n_1}\sqrt{n_1}(\widehat{\lambda}-\lambda^*)'\left[\frac{\partial^2}{\partial\lambda\partial\lambda'}\frac{1}{n_2}ssr^{\beta*}\right]\sqrt{n_1}(\widehat{\lambda}-\lambda^*)+o_p(1) \tag{25}$$

The RHS of (25) is $O_p(1)$ and what remains is to determine the joint distribution of all random quantities. Before tackling that issue, it is useful to examine the "misspecification / nonnestedness term" $W\equiv\left\{n_2\Delta+\frac{\sqrt{n_2}}{\sqrt{n_1}}\sqrt{n_2}\mu'\sqrt{n_1}(\widehat{\lambda}-\lambda^*)\right\}$, which consists of both components μ and Δ discussed in Remark 2. Clearly W is $O(n_2)$ as Δ is a constant. As mentioned in Remark 2 I assume it is negligible for practical purposes.

The term containing μ is more problematic as it contains a random quantity that also features on the RHS of (25). Additionally μ is needed in the rest of the analysis to calculate the variance of the first term on the RHS. Some manipulations yield:

$$\mu=E_{F_0}\left(\frac{\partial}{\partial\lambda}ssr_i^{\beta}(\beta^*,\lambda^*)\right)=\frac{\partial}{\partial\lambda}E_{F_0}\left(\varepsilon_{2i}+(z_i^0-x_i(\beta^*,\lambda^*))\right)^2$$

Denoting by E_K the expectation w.r.t. the distribution of the covariates (strikes) only one obtains (since $E_{F_0}(\varepsilon_{2i})=0$ and $\frac{\partial}{\partial\lambda}E_{F_0}(\varepsilon_{2i})^2=0$):

$$\mu=-2E_K\left(\frac{\partial}{\partial\lambda}x_i(\beta^*,\lambda^*)\right)(z_i^0-x_i(\beta^*,\lambda^*)) \tag{26}$$

and $z_i^0-x_i(\beta^*,\lambda^*)$ is small if the independence property holds and the SMN model is sufficiently flexible.

I now derive the distribution of all random components in (25). First apply a multivariate CLT to all quantities of the form $n_2^{-1/2}\frac{\partial}{\partial\beta}ssr^{\beta*}$ (analogously for γ and λ) to get

$$\begin{bmatrix} n_2^{-1/2}\frac{\partial}{\partial\gamma}ssr^{\gamma*} \\ n_2^{-1/2}\frac{\partial}{\partial\beta}ssr^{\beta*} \\ n_2^{-1/2}\left(\frac{\partial}{\partial\lambda}ssr^{\beta*}-n_2\mu\right) \\ n_1^{-1/2}\frac{\partial}{\partial\lambda}ssr^{\lambda*}\end{bmatrix}\sim N\left(0;\begin{bmatrix} B_{\gamma\gamma} & B_{\gamma\beta} & B_{\gamma\lambda} & 0 \\ B'_{\gamma\beta} & B_{\beta\beta} & B_{\beta\lambda} & 0 \\ B'_{\gamma\lambda} & B'_{\beta\lambda} & B_{\lambda\lambda} & 0 \\ 0 & 0 & 0 & G_{\lambda\lambda}\end{bmatrix}\right)$$

Here e.g. $B_{\gamma\beta}=E_{F_0}\left[\frac{1}{n_2}\left(\frac{\partial}{\partial\gamma}ssr^{\gamma*}\right)\left(\frac{\partial}{\partial\beta}ssr^{\beta*}\right)'\right]$ and $G_{\lambda\lambda}=E_{F_0}\left[\frac{1}{n_1}\left(\frac{\partial}{\partial\lambda}ssr^{\lambda*}\right)\left(\frac{\partial}{\partial\lambda}ssr^{\lambda*}\right)'\right]$ (G is used to avoid the clash of notation with the derivative of ssr^{β} w.r.t. λ). Denote the variance matrix in the above expression by B.

From Taylor expansions of the first-order conditions (i.e. the derivatives of the sum-of-squares functions) around the pseudo-true values one obtains the system of equations:

$$
\begin{aligned}
0 &= n_2^{-1/2}\frac{\partial}{\partial\gamma}ssr^{\gamma*} + \left[\frac{\partial^2}{\partial\gamma\partial\gamma'}\frac{1}{n_2}ssr^{\lambda*}\right]\sqrt{n_2}(\widehat{\gamma}-\gamma^*) + o_p(1) \\
0 &= n_2^{-1/2}\frac{\partial}{\partial\beta}ssr^{\beta*} + \left[\frac{\partial^2}{\partial\beta\partial\beta'}\frac{1}{n_2}ssr^{\beta*}\right]\sqrt{n_2}(\widehat{\beta}-\beta^*) \\
&\quad + \frac{\sqrt{n_2}}{\sqrt{n_1}}\left[\frac{\partial^2}{\partial\beta\partial\lambda'}\frac{1}{n_2}ssr^{\beta*}\right]\sqrt{n_1}(\widehat{\lambda}-\lambda^*) + o_p(1) \\
0 &= n_1^{-1/2}\frac{\partial}{\partial\lambda}ssr^{\lambda*} + \left[\frac{\partial^2}{\partial\lambda\partial\lambda'}\frac{1}{n_1}ssr^{\lambda*}\right]\sqrt{n_1}(\widehat{\lambda}-\lambda^*) + o_p(1)
\end{aligned}
$$

Denote the (scaled to be $O(1)$) second derivatives pertaining to the second equations in (15) and (16) by A, those for the first equation (i.e. related to λ) by H. E.g. $A_{\gamma\gamma} = \frac{\partial^2}{\partial\gamma\partial\gamma'}\frac{1}{n_2}ssr^{\lambda*}$ and so on. Rearrange the above to get

$$
\begin{aligned}
\sqrt{n_2}(\widehat{\gamma}-\gamma^*) &= -A_{\gamma\gamma}^{-1}\left(n_2^{-1/2}\frac{\partial}{\partial\gamma}ssr^{\gamma*}\right) + o_p(1) \\
\sqrt{n_2}(\widehat{\beta}-\beta^*) &= -A_{\beta\beta}^{-1}\left(n_2^{-1/2}\frac{\partial}{\partial\beta}ssr^{\beta*}\right) + \frac{\sqrt{n_2}}{\sqrt{n_1}}A_{\beta\beta}^{-1}\,A_{\beta\lambda}\,H_{\lambda\lambda}^{-1}\left(n_1^{-1/2}\frac{\partial}{\partial\lambda}ssr^{\lambda*}\right) + o_p(1) \\
\sqrt{n_1}(\widehat{\lambda}-\lambda^*) &= -H_{\lambda\lambda}^{-1}\left(n_1^{-1/2}\frac{\partial}{\partial\lambda}ssr^{\lambda*}\right) + o_p(1)
\end{aligned}
$$

where I have already substituted for λ in the second equation. All capital letters are $O(1)$ matrices and the expressions between () are asymptotically jointly normal as determined above. Hence one obtains the joint distribution of all 4 random quantities at the RHS of (25):

$$
\begin{bmatrix}
\sqrt{n_2}(\widehat{\gamma}-\gamma^*) \\
\sqrt{n_2}(\widehat{\beta}-\beta^*) \\
n_2^{-1/2}\left(\frac{\partial}{\partial\lambda}ssr^{\beta*} - n_2\mu\right) \\
\sqrt{n_1}(\widehat{\lambda}-\lambda^*)
\end{bmatrix}
= A.
\begin{bmatrix}
n_2^{-1/2}\frac{\partial}{\partial\gamma}ssr^{\gamma*} \\
n_2^{-1/2}\frac{\partial}{\partial\beta}ssr^{\beta*} \\
n_2^{-1/2}\left(\frac{\partial}{\partial\lambda}ssr^{\beta*} - n_2\mu\right) \\
n_1^{-1/2}\frac{\partial}{\partial\lambda}ssr^{\lambda*}
\end{bmatrix}
+ o_p(1)
$$

where

$$
A := \begin{bmatrix}
-A_{\gamma\gamma}^{-1} & 0 & 0 & 0 \\
0 & -A_{\beta\beta}^{-1} & 0 & \frac{\sqrt{n_2}}{\sqrt{n_1}}A_{\beta\beta}^{-1}\,A_{\beta\lambda}\,H_{\lambda\lambda}^{-1} \\
0 & 0 & I & 0 \\
0 & 0 & 0 & -H_{\lambda\lambda}^{-1}
\end{bmatrix}
$$

and I is the identity matrix of the appropriate dimension. Hence

$$
\begin{bmatrix}
\sqrt{n_2}(\widehat{\gamma}-\gamma^*) \\
\sqrt{n_2}(\widehat{\beta}-\beta^*) \\
n_2^{-1/2}\left(\frac{\partial}{\partial\lambda}ssr^{\beta*} - n_2\mu\right) \\
\sqrt{n_1}(\widehat{\lambda}-\lambda^*)
\end{bmatrix}
\sim N(0;\Sigma)
$$

with $\Sigma = ABA'$.

From Mathai and Provost (1992):

Definition 3 *(Weighted Sums of Chi-Square Distributions) Let $Z = (Z_1, ..., Z_m)'$ be a vector of m independent standard normal variables, and let $\omega = (\omega_1, ..., \omega_m)'$be a vector of m real numbers. Then, the random variable $\sum_{j=1}^{m} \omega_j Z_j^2$ is distributed as a weighted sum of chi-squares with parameters (m, ω). Its cumulative distribution function is denoted by $M_m(.;\omega)$.*

Lemma 1 *(distribution of quadratic forms) Let Y be a vector of m random variables distributed as N(0,Σ) with rank $\Sigma \leq m$. Let Ψ be a $m \times m$ real symmetric matrix. Then*

$$Y'\Psi Y \sim M_m(.;\omega)$$

where ω is the vector of eigenvalues of $\Psi\Sigma$. Moreover, the eigenvalues are all real, and they are all nonnegative if Ψ is positive semi-definite.

The RHS of expression (25) is a quadratic form with

$$\Psi = \begin{bmatrix} \frac{1}{2}A_{\gamma\gamma} & 0 & 0 & 0 \\ 0 & -\frac{1}{2}A_{\beta\beta} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2}\frac{\sqrt{n_2}}{\sqrt{n_1}}I \\ 0 & 0 & \frac{1}{2}\frac{\sqrt{n_2}}{\sqrt{n_1}}I & \frac{1}{2}\frac{n_2}{n_1}A_{\lambda\lambda} \end{bmatrix}$$

(I is again the identity matrix of the appropriate dimension). This finally gives the following simple version of Proposition 1:

Proposition 2 *The test-statistic $\widehat{ssr}^{\beta} - \widehat{ssr}^{\gamma}$ is approximately distributed as a weighted sum of chi-squares $M_m(.;\omega)$ with ω the eigenvalues of the matrix $\Psi\Sigma$ as defined above. $m = 2\dim(\lambda) + \dim(\gamma) + \dim(\beta)$.*

Note that the matrix $\Psi\Sigma$ is not known but can be consistently estimated. Based on this estimate, quantiles of the distribution can be computed by simulation.

I conclude by providing expressions for the constants Δ and μ_j in Proposition 1. Denote the estimated sum of squared residuals for T_2 options by $\widehat{ssr}^{I}$ for the case of estimation when imposing the independence property. Similarly, $\widehat{ssr}^{Aj}$ indicates the sum of squared residuals for T_2 options observed at time t_j estimated under the alternative. The sum of the latter over observation dates is denoted by $\widehat{ssr}^{A}$ (i.e. $\widehat{ssr}^{A} = \sum_{j=1}^{M} \widehat{ssr}^{Aj}$). When these functions are evaluated at the pseudo-true values of the parameters, a superscript * is appended. One then has

$$\mu_j := E_{F_0} \frac{\partial}{\partial\theta_{t_j,T_1}} ssr^{I*} \neq 0$$

with equality only under correct specification. Finally,

$$\Delta := \text{plim} \frac{1}{n_2}(ssr^{I*} - ssr^{A*})$$

where $n_2 := \sum_{j=1}^{M} n_{2j}$.

References

AHN, C. M. (1992): "Option Pricing When Jump Risk Is Systematic," *Mathematical Finance*, 2 (4), 299–308.

AÏT-SAHALIA, Y., AND A. W. LO (1998): "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *The Journal of Finance*, 53 (2), 499–547.

AMIN, K. I. (1993): "Jump Diffusion Option Valuation in Discrete Time," *Journal of Finance*, 48 (5), 1833–1863.

ANDERSEN, L., AND J. ANDREASEN (2000): "Jump-Diffusion Processes: Volatility Smile Fitting and Numerical Methods for Option Pricing," *Review of Derivatives Research*, 4, 231–262.

BAHRA, B. (1997): "Implied Risk-Neutral Probability Density Functions from Option Prices: Theory and Application," *working paper, Bank of England.*

BAKSHI, G., C. CAO, AND Z. CHEN (1997): "Empirical Performance of Alternative Option Pricing Models," *Journal of Finance*, 52, 2003–2049.

BAKSHI, G., C. CAO, AND Z. CHEN (2000): "Pricing and Hedging Long-Term Options," *Journal of Econometrics*, 94, 277–318.

BATES, D. S. (2000): "Post-'87 Crash Fears in the SP 500 Futures Option Market," *Journal of Econometrics*, 94, 181–238.

——— (2003): "Empirical option pricing: a retrospection," *Journal of Econometrics*, 116(1/2), 387–404.

BECKERS, S. (1980): "The constant elasticity of variance model and its implications for option pricing," *Journal of Finance*, 35(3), 661–673.

BLACK, F., AND M. SCHOLES (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–654.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

BURASCHI, A., AND J. JACKWERTH (2001): "The Price of a Smile: Hedging and Spanning in Option Markets," *Review of Financial Studies*, 14 (2), 495–527.

CARR, P., H. GEMAN, D. MADAN, AND M. YOR (2002): "The fine structure of asset returns: An empirical investigation," *Journal of Business*, 75, 305–332.

CARR, P., D. MADAN, H. GEMAN, AND M. YOR (2005): "From local volatility to local Lévy models," *Quantitative Finance*, forthcoming.

CARR, P., AND L. WU (2003): "The finite moment log stable process and option pricing," *Journal of Finance*, 58(2), 753–777.

——— (2004): "Time-changed Lévy processes and option pricing," *Journal of Financial Economics*, 71, 113–141.

Cartea, A., and S. Howison (2004): "Option pricing with Lévy-stable processes," *Working paper, Oxford Financial Research Centre*, 2004-MF-01, 27.

Clément, E., C. Gouriéroux, and A. Monfort (2000): "Econometric Specification of the Risk-Neutral Valuation Model," *Journal of Econometrics*, 94, 117–143.

Cox, J. C., and S. A. Ross (1976): "The Valuation of Options for Alternative Stochastic Processes," *Journal of Financial Economics*, 3, 145–166.

Cox, J. C., S. A. Ross, and M. Rubinstein (1979): "Option pricing: a simplified approach," *Journal of Financial Economics*, 7, 229–263.

De Wachter, S. (2003): "Evaluating the consistency of observed option prices with economic theory," Ph.D. thesis, Department of Economics, University of Oxford.

Delbaen, F., and H. Shirakawa (2003): "A note on option pricing for the constant elasticity of variance model," *mimeo*, p. 11.

Doornik, J. A. (2001): *Ox: An object-oriented matrix language*. Timberlake Consultants Press, London, 4th edn.

Dumas, B., J. Fleming, and R. E. Whaley (1998): "Implied Volatility Functions: Empirical Tests," *Journal of Finance*, 53, 2059–2106.

Dupire, B. (1994): "Pricing with a Smile," *Risk*, 7, 18–20.

Eberlein, E., U. Keller, and K. Prause (1998): "New insights into smile, mispricing, and value at risk: The hyperbolic model," *The Journal of Business*, 71(3), 371–405.

Engle, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation," *Econometrica*, 50, 987–1008.

Engle, R. F., and C. Mustafa (1992): "Implied ARCH Models from Options Prices," *Journal of Econometrics*, 52(1-2), 289–311.

Gourieroux, C., and A. Monfort (1995): "Testing, Encompassing and Simulating Dynamic Econometric Models," *Econometric Theory*, 11(2), 195–228.

Henderson, V., and D. Hobson (2003): "Coupling and Option Price Comparisons in a Jump-Diffusion Model," *Stochastics and Stochastic Reports*, forthcoming, 22p.

Heston, S. (1993a): "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bonds and Currency Options," *Review of Financial Studies*, 6, 327–343.

Heston, S. (1993b): "Invisible parameers in option prices," *Journal of Finance*, 48, 933–947.

Heston, S. L., and S. Nandi (2000): "A Closed-Form GARCH Option Valuation Model," *Review of Financial Studies*, 13(3), 585–625.

Hull, J. C., and A. White (1987): "The Pricing of Options with Stochastic Volatilities," *Journal of Finance*, 42, 281–300.

Jackson, N., E. Suli, and S. Howison (1999): "Computation of Deterministic Volatility Surfaces," *Journal of Computational Finance*, 2(2), 5–32.

Jacquier, E., and R. Jarrow (2000): "Bayesian Analysis of Contingent Claim Model Error," *Journal of Econometrics*, 94, 145–180.

Jondeau, E., and M. Rockinger (2000): "Reading the Smile: The Message Conveyed by Methods Which Infer Risk Neutral Densities," *Journal of International Money and Finance*, 19(6), 885–915.

Kou, S. (2002): "A jump diffusion model for option pricing," *Management Science*, 48, 1086–1101.

Kou, S., and H. Wang (2001): "Option Pricing Under a Double Exponential Jump Diffusion Model," *Working Paper, Columbia University.*

Madan, D. B., P. P. Carr, and E. C. Chang (1998): "The Variance Gamma Process and Option Pricing," *European Finance Review*, 2, 79–105.

Madan, D. B., and F. Milne (1991): "Option Pricing with VG Martingale Components," *Mathematical Finance*, 1 (4), 39–55.

Madan, D. B., and E. Seneta (1990): "The Variance Gamma (V.G.) Model for Share Market Returns," *Journal of Business*, 63 (4), 511–524.

Mathai, A. M., and S. B. Provost (1992): *Quadratic Forms in Random Variables: Theory and Applications*, vol. 126 of *Textbooks and Monographs in Statistics*. Marcel Dekker, New York, 1st edn.

McCullogh, J. H. (1996): "Financial applications of stable distributions," in *Statistical Methods in Finance*, ed. by G. S. Maddala, and C. R. Rao. North-Holland.

Melick, W., and C. Thomas (1997): "Recovering an Asset's Implied Pdf from Option Prices: An Application to Crude Oil During the Gulf Crisis," *Journal of Financial and Quantitative Analysis*, 32, 91–115.

Merton, R. C. (1976): "Option Pricing When Underlying Stock Returns are Discontinuous," *Journal of Financial Economics*, 3, 125–144.

Nicolato, E., and M. J. Venardos (2003): "Option Pricing in Stochastic Volatility Models of OU Type," *Mathematical Finance*, 13(4), 445–466.

Rebonato, R. (2000): *Volatility and Correlation in the Pricing of Equity, FX and Interest-Rate Options*, Financial Engineering. John Wiley and Sons, New York.

Ritchey, R. (1990): "Call Option Valuation for Discrete Normal Mixtures," *Journal of Financial Research*, 13, 285–295.

Rubinstein, M. (1994): "Implied Binomial Trees," *Journal of Finance*, 49 (3), 771–818.

Stein, E. M., and J. C. Stein (1991): "Stock Price Distributions with Stochastic Volatility: An Analytic Approach," *Review of Financial Studies*, 4, 727–752.

Taylor, S. (1986): *Modeling Financial Time Series*. Wiley and Sons, New York.

Vuong, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57, 307–333.

White, H. (1981): "Consequences and Detection of Misspecified Nonlinear Regression Models," *Journal of the American Statistical Association*, 76, 419–433.

Yatchew, A., and W. Härdle (2003): "Nonparametric state-price density estimation using constrained least squares and the bootstrap," *Journal of Econometrics*, forthcoming.

Zheng, J. X. (1996): "A consistent test of functional form via nonparametric estimation techniques," *Journal of Econometrics*, 75, 263–289.