

A SEMIPARAMETRIC ESTIMATION OF LIQUIDITY EFFECTS ON OPTION PRICING

Eva Ferreira
(Universidad del País Vasco)

Mónica Gago
(Universidad del País Vasco)

and

Gonzalo Rubio
(Universidad del País Vasco)

First draft: July 1999
This draft: September 1999

Keywords: Multivariate Kernel Regression, Symmetrized Nearest Neighbors, Bandwidth Selection, Volatility Smile, Risk-Neutral Densities, Liquidity, Option Pricing

Abstract

This paper proposes a semiparametric option pricing model with liquidity, as proxied by the relative bid-ask spread. The nonparametric volatility function with liquidity as an explanatory variable is estimated using the Symmetrized Nearest Neighbors (SNN) estimator rather than the traditional kernel estimator. Moreover, special care is taken in obtaining the smoothing parameter. The in-sample performance of the model turns out to be statistically favorable relative to a competing model without liquidity. However, the out-of-sample performance of both models is quite disappointing despite the fact that we are not able to reject the stability of risk-neutral densities estimated over different quarters during our sample period.

Corresponding author: Mónica Gago, Dpto. Econometría y Estadística, Facultad de Ciencias Económicas, Universidad del País Vasco, Avda. L. Aguirre 83, 48015 Bilbao, Spain; e-mail: etdgagam@bs.ehu.es

Eva Ferreira and Gonzalo Rubio acknowledge the financial support provided by Dirección Interministerial Científica y Técnica (DGICYT) grants PB95-0346 and PB97-0621 respectively. We appreciate the helpful comments of Ángel León.

1. Introduction

It is well understood that the central point for the empirical testing of option pricing models is whether the actual distribution of the underlying asset implied by the option market data is consistent with the distribution assumed by the theoretical option pricing model.

Given the Black-Scholes (1973) assumptions, all option prices on the same underlying security with the same expiration date but with different exercise prices should have the same implied volatility. However, the well known volatility smile pattern suggests that the BS formula tends to misprice deep in-the-money and deep out-of-the-money options¹. There have been various attempts to deal with this apparent failure of the BS valuation model. In principle, as explained by Das and Sundaram (1998) and others, the existence of the smile may be attributed to the well known presence of excess kurtosis in the conditional return distributions of the underlying assets. It is clear that excess kurtosis makes extreme observations more likely than in the BS case. This increases the value of out-of-the-money and in-the-money options relative to at-the-money options, creating the smile. However, at least in the U.S. market, the pattern shown by data contains a clear asymmetry in the shape of the smile. This may be due to the presence of skewness in the distribution which has the effect of accentuating just one side of the smile.

Given this evidence, extensions to the BS model that exhibit excess kurtosis and skewness have been proposed in recent years along two lines of research: Jump-diffusion models with a Poisson-driven jump process, and the stochastic volatility framework are the two key developments in the theoretical option pricing literature.

¹ After the October 1987 crash, the implied volatility computed from options on stock indexes in the US market inferred from the BS formula appears to be different across exercise prices. This is the so-called “volatility smile”. In fact, as pointed out by Rubinstein (1994), Ait-Sahalia and Lo (1998a) and Dumas, Fleming and Whaley (1998), implied volatilities of the S&P 500 options decrease monotonically as the exercise price becomes higher relative to the current level of the underlying asset. On the other hand, Taylor and Xu (1994) show that currency options tend to present a much more pronounced smile. Similar patterns of implied volatilities across exercise prices are found by Peña, Rubio and Serna (1999a) in the Spanish options market. Moreover, Bakshi, Cao and Chen (1997), and Fiorentini, León and Rubio (1998) report smile shapes for the (implied) instantaneous volatility under stochastic volatility and jump-diffusion option pricing models. León and Rubio (1999) theoretically study the behavior of the implied volatility function (smile) when the true distribution of the underlying asset is consistent with the stochastic volatility model proposed by Heston (1993).

Unfortunately, however, the empirical evidence regarding these new models is quite disappointing. Bates (1996), and Bakshi, Cao and Chen (1997) reject the jump-diffusion option pricing model on both currency options and equity options respectively. The stochastic volatility model proposed by Heston (1993) is rejected by Bakshi, Cao and Chen (1997), and Chernov and Ghysels (1998) for options written on S&P 500 index. Fiorentini, León and Rubio (1998) reject the same model for equity options on the Spanish IBEX-35 index.

These latter authors argue that the ultimate reasons behind the performance failure of Heston's model are closely related to the time-varying skewness and kurtosis found in the data. In particular, they suggest that the assumption of a constant correlation coefficient between returns and stochastic volatility should be relaxed if we really want to have a richer model. Unfortunately, the complexities needed to price options seem to increase without bounds. It may be the case that simple nonparametric (semiparametric) methodologies are able to incorporate the missing (realistic) factors in our option pricing models.

Along these lines, it should be pointed out that all previous models have been developed in a competitive, frictionless framework. It may be the case that liquidity costs, as represented by the percentage bid-ask spread, account for some of the differences observed between market prices and theoretical prices. Interestingly, Peña, Rubio and Serna (1999a) show that liquidity costs significantly cause the magnitude of the smile in equity options written on the Spanish IBEX-35 index².

A potentially relevant area of research might be related to endogenously incorporating liquidity costs in option pricing models with either stochastic volatility, stochastic jumps or both. A much more simple but, at the same time, effective approach would be based on the estimation of the implied volatility function with semiparametric methodologies, where the Black-Scholes implied volatility is replaced by a nonparametric function which depend upon a vector of explanatory variables. This is the multivariate kernel regression approach which has been recently followed by Aït-Sahalia and Lo (1998a). However, they ignore the potential effects of market frictions on the nonparametric volatility function. The objective of our paper is to fill up this gap

by incorporating, as an additional explanatory variable on the nonparametric volatility function, the percentage bid-ask spread. Thus, we construct the corresponding call pricing function under liquidity costs, and compare its performance relative to more traditional option pricing models. Hence, our nonparametric volatility function depends on moneyness, time to expiration and the percentage or relative bid-ask spread. In this sense, we are dealing with a multivariate nonparametric estimation³.

Moreover, we also estimate the *state price density* (SPD) or the so called (under non-arbitrage models) *risk-neutral density*, with the added potential effects of market frictions as proxied by the bid-ask spread. This is a key contribution of this paper to previous literature on option pricing. At the same time, and from a statistical point of view, our work improves the technique used by Aït-Sahalia and Lo (1998a) in, at least, two important ways: (i) The use of a multivariate kernel based on a global smoothing parameter may lead to estimation problems when obtaining the volatility nonparametric function in moneyness intervals for which the amount of data is relatively small. These intervals coincide with extreme out-the-money and in-the-money options and, of course, these are precisely the sections of the smile in which we are particularly interested. Given these arguments, we are planning to use the so called Symmetrized Nearest Neighbors (SNN) estimation instead of the more traditional kernel approach. (ii) Despite the fact that they have a three-dimensional kernel estimator, Aït-Sahalia and Lo employ a univariate smoothing parameter criterion. Moreover, they simplify the problem by eliminating the bias term in choosing the necessary bandwidth to estimate their nonparametric volatility function. We employ several criteria in order to calculate the bandwidth used in our estimations. In particular, we employ a plug-in criterion and multivariate approaches under three alternative specifications. Robustness relative to the bandwidth parameter is an important issue in nonparametric statistics.

² See also Longstaff (1995) and Dumas, Fleming and Whaley (1998).

³ Using a linear and a quadratic parametric approach, Peña, Rubio and Serna (1999b) solve numerically a forward partial differential equation with transaction costs also proxied by the relative bid-ask spread. Independently of the parametric specification, they find that these models seem to perform poorly relative to Black-Scholes.

We report both the in-sample and out-of-sample option pricing empirical results. We quarterly estimate our nonparametric multivariate volatility function and the corresponding option prices. Independently of the bandwidth criteria used, the in-sample results show an important improvement whenever we incorporate liquidity effects on the estimation. This result may have serious implications for option pricing research. However, the out-of-sample results are generally quite poor. All pricing models have a significant degree of mispricing. This should not be surprising taking into account that all models are estimated using quarterly data. This introduces a demanding requirement of stability to our estimated nonparametric functions, even though it should be noticed that, using a randomization test, we are not able to reject the stability of risk-neutral densities between quarters.

The paper is organized as follows. Section 2 contains a brief discussion on the relation between risk-neutral densities and non-arbitrage derivative pricing, and introduces our nonparametric estimation. In Section 3 we present the data available for our research. Section 4 discusses the nonparametric estimation of both the volatility smile and the risk-neutral density. The main empirical results are reported in Section 5. We first discuss the in-sample results, and secondly we present the stability tests of the risk-neutral densities over time and the out-of-sample performance of alternative pricing models. We conclude in Section 6 with a summary of results and a brief discussion of future work.

2. Nonparametric Estimation of Risk-Neutral Densities

2.1 Non-arbitrage Pricing

It is well known that, under risk-neutrality, the price of any financial asset can be expressed as the expected present value of its future payoffs, where the present value is obtained relative to the riskless rate and the expectation is taken relative to the risk-neutral density function of the payoffs.

For completeness⁴, we provide a brief summary of this non-arbitrage framework. Let $W_t = (W_{1t}, \dots, W_{dt})$ be a d -dimensional Brownian motion. Let the stock prices be represented by the following general model:

$$dS_{jt} = \mu_{jt} S_{jt} dt + S_{jt} \sum_{k=1}^d \sigma_{jkt} dW_{kt} \quad , \quad j = 1, \dots, m \quad (1)$$

Let us consider next the so called *market price of risk* equations:

$$\sum_{k=1}^d \sigma_{jkt} \theta_{kt} = \mu_{jt} - r_t \quad , \quad j = 1, \dots, m \quad (2)$$

where m is the number of equations (number of stocks), d is the number of unknowns (number of sources of randomness), and θ_t is the market price of risk.

Moreover, we can use the Martingale Representation Theorem⁵ to guarantee the existence of a d -dimensional process $\Psi_t = (\Psi_{1t}, \dots, \Psi_{dt})$ so that we may solve the so called *hedging equations* for $\Delta_1, \dots, \Delta_m$:

$$\sum_{j=1}^m \Delta_{jt} S_{jt} \sigma_{jkt} = \rho_t \Psi_{kt} \quad , \quad k = 1, \dots, d \quad (3)$$

where m is the number of unknowns (number of stocks), d is the number of equations (number of sources of randomness), and $\rho_t = \exp\left\{\int_0^t r_u du\right\}$.

⁴ See also Ait-Sahalia and Lo (1998a, 1998b).

⁵ See Shreve, Chalasani and Jha (1997).

We say that the market is complete if and only if the system (2) has a unique solution. Then, this implies that (3) always has a solution, and every contingent claim can be hedged. Finally, if $\theta_t = (\theta_{1t}, \dots, \theta_{dt})$ is the unique solution to (2), we define,

$$Z_t = \exp \left\{ -\int_0^t \theta_u dW_u - \frac{1}{2} \int_0^t \|\theta_u\|^2 du \right\}$$

$$P^*(A) = \int_A Z_T dP, \quad \forall A \in \Omega$$

where Z_t has finite variance. This probability measure, P^* , is the unique risk-neutral measure of the model.

In this case, the risk-neutral density or SPD can be characterized without any explicit reference to preferences. When we assume a geometric Brownian motion with constant volatility and interest rate, the risk-neutral density is given by the conditional distribution of the risk-neutral stochastic process given by,

$$dS_t = rS_t dt + \sigma S_t dW_t^*$$

which is the well known lognormal distribution with mean $(r - \sigma^2/2)(T-t)$ and variance $\sigma^2(T-t)$.

In general, if S_t is the stock price and $f_t^*(S_T)$ is the date-t risk-neutral density of the stock price at future date T, then any european derivative with a payoff at expiration given by some function $\varphi(S_T)$ can be priced by the following expression,

$$e^{-r(T-t)} E^*[\varphi(S_T)] = e^{-r(T-t)} \int_0^\infty \varphi(S_T) f_t^*(S_T) dS_T \quad (4)$$

where r is the constant riskless rate between t and T .

When the derivative is a european call option with expiration at T and exercise price, X , its price would be:

$$c = e^{-r(T-t)} \int_0^{\infty} \max[S_T - K] f_t^*(S_T) dS_T \quad (5)$$

Finally,

$$f_t^*(S_T) = e^{r(T-t)} \left. \frac{\partial^2 c}{\partial X^2} \right|_{X=S_T} \quad (6)$$

so that, the risk-neutral density is proportional to the second derivative of the option price function with respect to the exercise price.

2.2 Nonparametric Estimation

As discussed in the introduction, the idea of the paper is to estimate the risk-neutral density nonparametrically, and to be able to price options. Our procedure is based on the following sequence of estimations: We employ option market prices to estimate a nonparametric volatility function which depends upon the degree of moneyness, time to expiration and liquidity, proxied by the relative bid-ask spread. Then, given this function in which volatility is allowed to vary with moneyness, time to expiration and the bid-ask spread, the Black-Scholes formula can be used to obtain, semiparametrically, option prices⁶. In the last step, we differentiate this option estimator twice with respect to the exercise price to obtain (6), given the appropriate interest rate. The issue, of course, is how to estimate the multivariate volatility function nonparametrically.

⁶ This semiparametric estimation of option prices considerably reduces the dimensionality of the problem. As pointed out by Aït-Sahalia and Lo (1998a), the sample size required to achieve the same degree of accuracy as in the full nonparametric estimation may be much smaller.

It is important to realize that there is no an obvious way to model the influence of moneyness, time to expiration and liquidity on the volatility function. It is precisely in this sense in which the nonparametric framework provides a very flexible approach.

Let us consider a multivariate kernel estimator of the volatility function, with possibly different smoothing parameters for the covariates:

$$\tilde{\sigma}(\xi, SP, \tau) = \frac{\frac{1}{nh_{\xi}h_{SP}h_{\tau}} \sum_{i=1}^n K\left(\frac{\xi - \xi_i}{h_{\xi}}\right) K\left(\frac{SP - SP_i}{h_{SP}}\right) K\left(\frac{\tau - \tau_i}{h_{\tau}}\right) \sigma_i}{\frac{1}{nh_{\xi}h_{SP}h_{\tau}} \sum_{i=1}^n K\left(\frac{\xi - \xi_i}{h_{\xi}}\right) K\left(\frac{SP - SP_i}{h_{SP}}\right) K\left(\frac{\tau - \tau_i}{h_{\tau}}\right)} \quad (7)$$

where $\xi \equiv X/F$ is the degree of moneyness, where X is the exercise price and F is the futures price (underlying asset)⁷, SP is the relative bid-ask spread, τ is the time to expiration, σ_i is the volatility implied by the option price c_i , h_j is the bandwidth or smoothing parameter for each covariate $j = \xi, SP, \tau$, and $\tilde{\sigma}(\xi, SP, \tau)$ is the three-dimensional nonparametric volatility function to be estimated.

It is important to point out that a kernel estimator based on a global smoothing parameter may lead to poor estimation results basically in those zones where we have a relatively small amount of data. Translating these effects to our case, it suggests that we may obtain poor estimations for the volatility function for out-of-the-money and in-the-money options. Of course, from a financial point of view, these are precisely the options (and the sections of the volatility smile) which we are particularly interested in.

Given this fact, in this paper we employ the Symmetrized Nearest Neighbors (SNN) estimation as an alternative to the classical kernel estimator. This kind of estimators was proposed by Yang (1981), and studied in detail by Stute (1984). The idea behind them is

⁷ Note that the underlying asset in the Spanish market for which we have data is the futures price on the stock exchange index.

very simple. When estimating in one point we calculate the weight for the rest of observations looking at the distance between the values of the empirical distribution at each point rather than the distance between the points themselves. Hence, the estimator is defined as:

$$\hat{\sigma}(\xi, SP, \tau) = \frac{\frac{1}{nh_\xi h_{SP} h_\tau} \sum_{i=1}^n K\left(\frac{F_n(\xi) - F_n(\xi_i)}{h_\xi}\right) K\left(\frac{F_n(SP) - F_n(SP_i)}{h_{SP}}\right) K\left(\frac{F_n(\tau) - F_n(\tau_i)}{h_\tau}\right) \sigma_i}{\frac{1}{nh_\xi h_{SP} h_\tau} \sum_{i=1}^n K\left(\frac{F_n(\xi) - F_n(\xi_i)}{h_\xi}\right) K\left(\frac{F_n(SP) - F_n(SP_i)}{h_{SP}}\right) K\left(\frac{F_n(\tau) - F_n(\tau_i)}{h_\tau}\right)} \quad (8)$$

where $F_n(\cdot)$ denotes the empirical distribution of corresponding variable, ξ , SP or τ .

Roughly speaking, the empirical distribution changes the random design to a uniform design with the knots uniformly spaced between zero and one. In practice, using SNN estimators is basically the same as employing kernel estimators for $F_n(X_i)$ instead of X_i . A detailed discussion on the differences between these estimators is contained in Appendix A, where we present and compare the minimum asymptotic mean square error (MSE) for both kernels and SNN. Moreover, to provide some intuition related to our particular case, we discuss an example that illustrates the behavior of our dataset.

As shown in Appendix A, both estimators have the same MSE under a uniform design. On the other hand, if we assume the bias to be negligible with respect to variance, it is easy to show that using the SNN estimator with bandwidth h is equivalent to employ a kernel estimator with variable bandwidth equal to $h/f(x)$. Also, the discussion provided in Appendix A allows us to argue that, in the tails, a smaller MSE is obtained for the SNN estimator. It should be recalled that we are particularly concerned with the tails of the distribution given, of course, that extreme degrees of moneyness are a key issue in terms of both the volatility smile and pricing.

Once we have estimated the volatility function given by (8), we have to estimate the call-pricing function. This function is evaluated as, $\hat{c}(\xi, r, \tau, \hat{\sigma}(\xi, SP, \tau))$, where the

function $\hat{c}(\cdot)$ is the same as in the Black-Scholes expression with the nonparametrically estimated volatility. This is to say,

$$\hat{c}(\xi, \tau, r, \hat{\sigma}(\xi, SP, \tau)) = c_{BS}(\xi, \tau, r, \hat{\sigma}(\xi, SP, \tau)) \quad (9)$$

The risk-neutral density estimator follows by taking the appropriate partial second derivative of $\hat{c}(\cdot)$ with respect to the exercise price. The detailed derivation of this second derivative is reported in Appendix B:

$$\hat{f}_t^*(S_T) = e^{r(T-t)} \left[\frac{\partial^2 \hat{c}(\xi, \tau, r, \hat{\sigma}(\xi, SP, \tau))}{\partial X^2} \right]_{X=S_T} \quad (10)$$

In practice, the last and probably most important problem faced up by any researcher is the selection of the smoothing parameters, (h_ξ, h_{SP}, h_τ) . It is interesting to point out that there is a tremendous amount of literature developed for the univariate case. See Härdle (1990) for a general presentation of this literature. Unfortunately, the bandwidth selection becomes much more complicated in the multivariate context.

Generally speaking, there are two groups of methods to select the bandwidth: plug-in methods and methods based on the minimization of some penalized least square error measure. To decide the particular selection method to be employed is not a trivial task. Even in the simpler case, in which we have a fixed design and an univariate estimator, different asymptotically optimal methods may lead to different smoothing parameters. In our case, we have not only a random design but also a multivariate context.

For a similar context as ours, Aït-Sahalia and Lo (1998a) use a global univariate selection criterion. However, this criterion do not take into account the multivariate character of the estimator (the optimum univariate bandwidths might be different from the optimal bandwidths in the multivariate context), and, moreover, they do not offer any criteria to choose the constant involved in the estimation.

It seems clear to us that the bandwidth parameter is the most important quantity to be selected for any nonparametric estimation; it must definitely be carefully selected. It seems, therefore, convenient to analyze the stability of the smoothing estimators, and the robustness of results, relative to alternative selection methodologies. Appendix C contains a detailed discussion of the alternative techniques employed in this paper to calculate the bandwidths.

With these considerations in mind, and being concerned with computational costs whenever a very sophisticated method is used, we propose the following methodology. We first compute the univariate *pilot* bandwidths for our three explanatory variables by using a plug-in method, where the constants are selected with an iterative method due to Gasser, Kneip and Köhler (1991) and discussed in Appendix C⁸. Since the asymptotical rates of convergence suggest that the bandwidth is influenced by the dimensionality of the problem, we check the validation of these *pilot* bandwidths by using three alternative multivariate cross-validation criteria. In particular, we employ the natural extensions of Generalized Cross-Validation (GCV) method and Rice's bandwidth selectors to the multivariate case. All of them are presented in Appendix C. They are evaluated in a grid of bandwidths around the *pilot* smoothing parameters.

As it will be shown later when presenting the empirical results, it is observed that the multivariate criteria tend to select slightly higher parameters. There is some intuition behind this result. The rate at which h must go to zero is of order $n^{-1/(4+d)}$, where d denotes the dimension of the covariate vector. Therefore, it should not be surprising to see that, whenever we employ a multivariate criterion, h becomes larger. On the other hand, given that the final estimators do not change substantially, we report most of our results on the basis of the plug-in smoothing parameters selector. More will be said later on these issues.

⁸ We are currently working in extending this iterative method to our multivariate case.

3. The Data: The Spanish IBEX-35 and the Option Contract

The Spanish IBEX-35 index is a value-weighted index comprising the 35 most liquid Spanish stocks traded in the continuous auction market system. The official derivative market for risky assets, which is known as MEFF, trades a futures contract on the IBEX-35, the corresponding option on the IBEX-35 futures contracts for calls and puts, and individual option contracts for blue-chip stocks. Trading in the derivative market started in 1992. The market has experienced tremendous growth from the very beginning. Relative to the volume traded in the Spanish continuous market, trading in MEFF represented 40% of the regular continuous market in 1992 and 138% in 1996. The number of all traded contracts in MEFF relative to the contracts traded in the CBOE reached 21% in 1996.

The Spanish option contract on the IBEX-35 futures is a cash settled European option with trading during the three nearest consecutive months and the other three months of the March-June-September-December cycle. The expiration day is the third Friday of the contract month. Trading occurs from 10:30 to 17:15. During the sample period covered by this research, the multiplier has changed from 100 Spanish pesetas times the IBEX-35 index at the beginning of the sample period to 1000 pesetas during 1998, and prices are quoted in full points, with a minimum price change of one index point⁹. The exercise prices are given by 50 index point intervals.

It is important to point out that liquidity is concentrated in the nearest expiration contract. In fact, during the sample period almost 90% of crossing transactions occurred in this type of contracts.

Our database is comprised of all call and put options on the IBEX-35 index futures traded daily on MEFF during the period January 1996 through November 1998. Given the concentration in liquidity, our daily set of observations includes only calls and puts with two possible expiration dates. We only include options which expire between five and forty days. That is, we eliminate all transactions taking place during the last five

⁹ Starting in January 1999, it has been changed to 10 euros.

days before expiration, and transactions which will expire in more than forty days.

As usual in this type of research, our primary concern is the use of simultaneous prices for the options and the underlying security. The data, which are based on all reported transactions during each day throughout the sample period, do not allow us to observe simultaneously enough options with the same time-to-expiration on exactly the same underlying security price but with different exercise prices. In order to avoid large variations in the underlying security price, we restrict our attention to the 45- minute window from 16:00 to 16:45. It turns out that almost 25% of crossing transactions occur during this interval. Moreover, care was also taken to eliminate the potential problems with artificial trading that are most likely to occur at the end of the day. Thus, all trades after 16:45 were eliminated so that we avoid data which may reflect trades to influence market maker margin requirements. At the same time, using data from the same period each day avoids the possibility of intraday effects in the IBEX-35 futures options market. Finally, we eliminate all call and put prices that violate the well known arbitrage bounds. The number of observations within a day may vary according to the number of crossing transactions associated with different exercise prices available each day.

These exclusionary criteria yield a final daily sample of 8321 observations (4798 calls and 3523 puts). The implied volatility for each of our 8321 options is estimated next. To do so, note that we take as the underlying asset the average of the bid and ask price quotation given for each futures contract associated with each option during the 45-minute interval. To proxy for riskless interest rates, we use the daily series of annualized repo T-bill rates with either one week, two weeks or three weeks to maturity. One of these three interest rates will be employed depending upon how close the option is to the expiration day.

Table 1 describes the sample properties of the call and put option prices employed in this work. Average prices, average relative bid-ask spread and the number of available calls and puts are reported for each moneyness category. Moneyness is defined as the ratio of the exercise price to futures price. A call (put) option is said to be deep out-of-the-money (deep in-the-money) if the ratio K/F belongs to the interval $(1.03, 1.08)$; out-of-the-money (in-the-money) if $1.03 > K/F \geq 1.01$; at-the-money when $1.01 > K/F \geq$

0.99; in-the-money (out-of-the-money) when $0.99 > K/F \geq 0.97$; and deep-in-the-money (deep out-of-the-money) if $0.97 > K/F \geq 0.90$. As we already discussed, there are 4798 call option observations (3523 puts), with OTM, ATM and ITM call (put) options respectively representing 61% (68), 30% (25) and 9% (7). The average call (put) price ranges from 61.6 (64.1) pesetas for deep OTM call (put) options to 381.2 (461.7) pesetas for deep ITM call (put) options. The average relative bid-ask spread tends to move in the opposite direction to the average price. In particular, it ranges from 0.38 (0.33) for deep OTM call (put) options to 0.12 (0.12) for deep ITM calls (ITM puts).

4. Smiles and Risk-neutral Densities

It is well recognized that option prices provide the market participants with a tremendous amount of information. In particular, we have already discussed how to infer the risk-neutral density or, alternatively, the Arrow-Debreu prices from trading options on the market portfolio. This section presents our nonparametric estimation of both the risk-neutral density and the volatility smile for 1998¹⁰. Implications for the behavior of the market portfolio, as represented by the IBEX-35 stock exchange index, are drawn on the basis of the implied (risk-neutral) distribution embedded in option prices.

We first discuss the nonparametric estimation of the univariate (traditional) volatility smile. As before, we employ the SNN estimator instead of the classical kernel estimator, and the plug-in bandwidth selection method:

$$\hat{\sigma}(\xi) = \frac{\frac{1}{nh_{\xi}} \sum_{i=1}^n K\left(\frac{F_n(\xi) - F_n(\xi_i)}{h_{\xi}}\right) \sigma_i}{\frac{1}{nh_{\xi}} \sum_{i=1}^n K\left(\frac{F_n(\xi) - F_n(\xi_i)}{h_{\xi}}\right)} \quad (11)$$

Expression (11) is firstly estimated using only call options transacted during 1998. It is interesting to note the important differences obtained when we employ the SNN

¹⁰ During 1998, financial markets experimented an enormous amount of volatility. Our presentation for 1998 should be just taken as a working example. In any case, the main implications of our analysis could have been obtained with any other year of the sample period.

procedure rather than the traditional kernel estimator. Figure 1 contains the nonparametric volatility smile estimated with both methods. It turns out that the optimal bandwidth under the SNN estimator is 0.093 and, as we can see in Figure 1, the smile is optimally smoothed. A very different picture arises when we employ the traditional kernel estimator given by:

$$\tilde{\sigma}(\xi) = \frac{\frac{1}{nh_\xi} \sum_{i=1}^n K\left(\frac{\xi - \xi_i}{h_\xi}\right) \sigma_i}{\frac{1}{nh_\xi} \sum_{i=1}^n K\left(\frac{\xi - \xi_i}{h_\xi}\right)} \quad (12)$$

The effect of undersmoothing is clearly reflected in Figure 1. The estimator is highly variable because the optimal bandwidth selected according to the plug-in method tends to zero. We argue that, given the specific data we usually have when doing research in option pricing, where a lot of observations are centered around the at-the-money options and relatively few observations are available in the extremes, the SNN estimator is more appropriate.

It is also well known that, given equation (10), the pattern of implied volatilities for alternative exercise prices (the smile) gives us a direct evidence of the risk-neutral density function actually embedded in option price data. Note that Figure 1 suggests that implied volatilities are lower for low-exercise-price options than for high-exercise price options. This is consistent with a distribution having a fatter right tail and a thinner left tail relative to the normal distribution. In other words, Figure 1 seems to imply that, during 1998, a positive relationship between volatility shocks and price changes of the underlying asset was actually the case in the Spanish market.

To further investigate this issue, we run both a Nagarch (1,1) and a GJR (1,1)¹¹ models in which we are allowed to calculate the correlation between conditional variance and the returns of the underlying asset. Let R_t be the return generated by the following model:

$$R_t = \mu + \varepsilon_t$$

¹¹ GJR refers to Glosten, Jagannathan, and Runkle (1993).

where,

$$\varepsilon_t \approx N(0, V_t)$$

$$\varepsilon_t = V_t^{1/2} \eta_t ; \eta_t \approx \text{iid } N(0,1)$$

Under the Nagarch(1,1), the conditional variance, V_t , and the corresponding correlation coefficient are defined as:

$$V_t = \omega + \beta V_{t-1} + \alpha \left(\varepsilon_{t-1} + \gamma \sqrt{V_{t-1}} \right)^2 \quad (13)$$

where γ represents the relation between the shocks and the conditional variance, and $\omega, \beta, \alpha > 0$, and where the correlation is:

$$\rho = \frac{\sqrt{2\gamma}}{\sqrt{1 + 2\gamma^2}}$$

Under the GJR(1,1) model we have that:

$$V_t = \omega + \beta V_{t-1} + (\alpha + \gamma D_{t-1}) \varepsilon_{t-1}^2 \quad (14)$$

where D_t is a dummy variable defined as:

$$D_t \equiv \begin{cases} 1 & \text{if } \varepsilon_t < 0 \\ 0 & \text{if } \varepsilon_t > 0 \end{cases}$$

and the correlation coefficient is¹²:

$$\rho = \frac{-\gamma}{\sqrt{\frac{\pi}{2} \left(2\alpha^2 + 2\alpha\gamma + 5/4 \gamma^2 \right)}}$$

The actual estimations using daily returns for the IBEX-35 stock exchange index during 1998 result in a significant and negative correlations of -0.753 and -0.701 for the Nagarch(1,1) and GJR(1,1) respectively¹³.

¹² Note that in this model a positive γ implies a negative correlation coefficient.

The explanation about this apparent contradiction between the direct evidence provided by the underlying asset, and the implicit results obtained from option prices lies on the lack of call transactions available for in-the-money calls. This suggest that, if we want to extract information about the behavior of the underlying asset, and correctly estimate the risk-neutral density of the underlying asset implied by option prices, we should use both calls and puts. This argument is not related to the put-call parity relationship. It is simply a consequence of lack of transactions in a relatively thin option market.

Figure 2 contains the nonparametric estimation of the volatility smile using both calls and puts. As we can observe, once the full option market is incorporated into the analysis, a typical asymmetric smile is obtained. This volatility function is consistent with a negative asymmetric distribution of the underlying asset.

We next estimate the implied risk-neutral distribution recognizing the potential effects of, not only moneyness, but also time to expiration and liquidity.

As we already mentioned in Section 2, we first estimate the nonparametric function, $\hat{\sigma}(\xi, SP, \tau)$ where ξ is the degree of moneyness, SP is the relative bid-ask spread, and τ is time to expiration. We employ the SNN estimator given by expression (8). It should be pointed out that we also estimate the nonparametric volatility function without the liquidity variable. This is an important issue in this paper. These estimations will allow us to compare the implied risk-neutral densities with and without liquidity effects.

Table 2 contains the optimal bandwidths given by the plug-in criterion for our three explanatory variables and for the three years in our sample. Given the high degree of uncertainty experimented by the market during 1998, it should not be surprising to observe that the bandwidths become larger for all variables during 1998. Otherwise, the results seem to be reasonable.

Once we have the nonparametric volatility functions with and without liquidity, we employ equation (10) to estimate the risk-neutral density function for 1998. In order to

¹³ The correlations for the whole sample period (1996-1998) are also negative and significant. They are

do so, we first observe all exercise prices available during 1998. The rest of the explanatory variables are assumed to be constant in their means for that year, so that the future price, the relative spread and time to expiration remain constant in their means, $\bar{F}, \bar{SP}, \bar{\tau}$. Note that the only variable allowed to vary is the exercise price. We now estimate the nonparametric volatility function given by (8) in the new knots $(X_i/\bar{F}, \bar{SP}, \bar{\tau})$ where $X_i, i = 1, \dots, n$, are the number of exercise prices observed for that particular period. The implied risk-neutral distribution is then given by¹⁴:

$$\hat{f}_t^*(X_i) = e^{r(T-t)} \left[\frac{\partial^2 \hat{c}(X_i/\bar{F}, \bar{\tau}, \bar{r}, \hat{\sigma}(X_i/\bar{F}, \bar{SP}, \bar{\tau}))}{\partial X^2} \right]_{X_i=S_T} \quad (15)$$

Since we are interested in the effects of liquidity on option pricing, the same procedure is repeated without taking into account the bid-ask spread variable. Thus, we have two risk-neutral densities, where in the first one the potential effects of market frictions are explicitly considered. As a reference, we also present the Black-Scholes distribution by employing the at-the-money mean implied volatility during 1998 as an input in the expression of the lognormal density. Note that this (mean) implied volatility is used as a constant volatility in all knots where the density is estimated.

Figure 3 contains the estimated risk-neutral density for 1998 in terms of returns rather than levels of exercise prices. This figure is obtained using all calls and puts available in the sample. The returns are simply calculated as $R_i = \ln(X_i/\bar{F})$. The results suggest a (slightly) negatively skewed distribution relative to Black-Scholes¹⁵, where the density estimated without liquidity presents a fatter left tail relative to the density estimated with liquidity, and a thinner right tail. This implies that the model without liquidity would underprice out-of-the-money puts and in-the-money calls relative to the model with liquidity, and, at the same time, the model without liquidity would overprice out-of-the-money calls and in-the-money puts relative to the option model incorporating liquidity.

equal to -0.407 and -0.310 for the same models.

¹⁴ In fact, the estimation of the implied risk-neutral density is easily simplified by noting that the only relevant term of equation (B.2), in terms of magnitude, is the first component of the expression. The last two terms are very small since they are multiplied by $1/X$ and $1/F^2$ respectively.

This evidence suggests that there might be relevant effects of liquidity on option pricing. They will be further investigated in the next section. It should also be noted that the implied distributions estimated nonparametrically are clearly leptokurtic relative to Black-Scholes.

Finally, Figures 4 and 5 present the densities estimated with either only calls and only puts respectively. Again, it is clear that one should be very careful in driving conclusions about the behavior of the underlying asset without considering simultaneously calls and puts in the estimation. This might be particularly important in relatively thin option markets, where trading is much more concentrated than in markets like the US option market.

5. Empirical Results

5.1 In-sample pricing performance

This section evaluates the performance of the alternative semiparametric option pricing models described above.

Two theoretical option prices are calculated using the nonparametric volatility functions estimated according to our two versions of equation (8) with and without liquidity. Once we have these functions, we employ Black's (1976) model to obtain our theoretical semiparametric option prices for each call in the sample. In particular, we calculate the following call prices by:

$$\hat{c}_i(\xi_i, \tau_i, r_i, \hat{\sigma}_i) = e^{-r_i(T-t)_i} [F_i N(d_{1i}) - X_i N(d_{2i})] \quad (16)$$

where,

$$d_{1i} = \frac{\ln(F_i/X_i) + \hat{\sigma}_i/2(T-t)_i}{\hat{\sigma}_i\sqrt{(T-t)_i}}, \quad d_{2i} = d_{1i} - \hat{\sigma}_i\sqrt{(T-t)_i} \quad (17)$$

where each option i available in the sample is characterized by a futures prices, F_i , an exercise price, X_i , a time to expiration, $(T-t)_i$ and, given the days to maturity, the

¹⁵ This is consistent with the volatility smile shown in Figure 2.

corresponding repo rate with similar maturity, r_i . As the input for volatility, $\hat{\sigma}_i$, we introduce the estimated (nonparametrically) volatility function which is obtained by either one of the following two SNN estimators:

$$\hat{\sigma}_{WL} = \frac{\frac{1}{nh_\xi h_{SP} h_\tau} \sum_{i=1}^n K\left(\frac{F_n(\xi) - F_n(\xi_i)}{h_\xi}\right) K\left(\frac{F_n(SP) - F_n(SP_i)}{h_{SP}}\right) K\left(\frac{F_n(\tau) - F_n(\tau_i)}{h_\tau}\right) \sigma_i}{\frac{1}{nh_\xi h_{SP} h_\tau} \sum_{i=1}^n K\left(\frac{F_n(\xi) - F_n(\xi_i)}{h_\xi}\right) K\left(\frac{F_n(SP) - F_n(SP_i)}{h_{SP}}\right) K\left(\frac{F_n(\tau) - F_n(\tau_i)}{h_\tau}\right)} \quad (18)$$

$$\hat{\sigma}_{WOL} = \frac{\frac{1}{nh_\xi h_\tau} \sum_{i=1}^n K\left(\frac{F_n(\xi) - F_n(\xi_i)}{h_\xi}\right) K\left(\frac{F_n(\tau) - F_n(\tau_i)}{h_\tau}\right) \sigma_i}{\frac{1}{nh_\xi h_\tau} \sum_{i=1}^n K\left(\frac{F_n(\xi) - F_n(\xi_i)}{h_\xi}\right) K\left(\frac{F_n(\tau) - F_n(\tau_i)}{h_\tau}\right)} \quad (19)$$

It should be pointed out that these two equations are estimated quarterly over the whole sample period from January 1996 to November 1998. The plug-in criterion is employed to calculate the optimal bandwidth for each explanatory variable and for each quarter. In this way, we have 4798 pricing errors for calls from January 2, 1996 to November 10, 1998, and for each of the models analyzed.

Table 3 reports two measures of performance for the alternative model specifications. Panel A contains the absolute pricing error which is the sample average of the squared difference between the model price and the market price for each call in the sample period. In Panel B, the reported percentage pricing error is the squared sample average of the theoretical price minus the market price divided by the market price. These two statistics are calculated for each moneyness categories and for all calls in the sample. They are given by the following expressions:

$$\text{APE} = \sqrt{\frac{1}{N_j} \sum_{i=1}^{N_j} (c_i^{\text{TH}} - c_i^{\text{M}})^2} \quad (20)$$

$$\text{PPE} = \sqrt{\frac{1}{N_j} \sum_{i=1}^{N_j} \left(\frac{c_i^{\text{TH}} - c_i^{\text{M}}}{c_i^{\text{M}}} \right)^2}$$

where c_i^{TH} denotes the estimated price for the call option using a particular option pricing model, and c_i^{M} is the corresponding option price observed in the market. N represents the number of calls in each moneyness category ($j = \text{Deep OTM, OTM, } \dots$)

The results of Table 3 are reported by years. The evidence is quite striking. In all cases, both the APE and PPE are lower when liquidity is incorporated in the estimation of the nonparametric volatility function. It is clear that pricing errors for our nonparametric model with liquidity are larger during 1998, but it is also the case that they are even larger whenever we do not price options taking into account the bid-ask spread.

In this paper, the statistical significance of performance is assessed by analyzing the proportion of theoretical prices lying outside their corresponding bid-ask spread boundaries. This will allow us to test whether or not the differences between the pricing performance of our two competing models are statistically different from zero.

For each model and each moneyness category, we compute the proportion of options such that the estimated theoretical price falls outside the bid-ask boundary. Let us denote by p_1 the proportion of prices outside the boundary when the model does not include liquidity, and p_2 the proportion when the bid-ask spread is taken into account. We want to test whether $p_2 < p_1$. If this were the case, we may argue that liquidity is a relevant variable in pricing call options.

One should be careful in defining this statistic. It should be noted that these two models are not independent. In fact, they are nested models. Thus, if we consider the difference $\hat{p}_2 - \hat{p}_1$, where \hat{p}_2 and \hat{p}_1 denote the estimated proportions, the asymptotic normal distribution under the null does not have the usual variance term. The reason is that the

proportions are not independent (in fact, they are directly dependent) and the variance of the difference should be smaller than the variance for the independent case.

To avoid this problem, we compute \hat{p}_1 and we employ this proportion, which is the one for the most restrictive case, as our null hypothesis for the second model. In other words, we test $H_0: p_2 \geq \hat{p}_1$ against $H_a: p_2 < \hat{p}_1$. In this context, the Z-statistic is given by:

$$Z = \frac{\hat{p}_2 - \hat{p}_1}{\sqrt{\hat{p}_1(1 - \hat{p}_1)/n}} \quad (21)$$

where n is the sample size. This statistic is asymptotically distributed as a standardized normal variable.

Given that we are also interested in knowing whether a given theoretical valuation model undervalues or overvalues market prices, the Z-statistic is also calculated to obtain the proportion for which the theoretical model yields a price below the bid quote, and the proportion for which the model gives a price above the ask quote. If a theoretical model tends to undervalue market prices, it would yield a higher proportion of prices below the bid quote. If, on the other hand, the model tends to overvalue market prices, it would have a higher proportion of prices above the ask quote.

Table 4 contains the results for each year and for each moneyness category. As in the previous table, the empirical results are quite impressive. For OTM and ATM calls the pricing performance of the semiparametric option pricing model with liquidity is statistically and systematically superior to the model without liquidity. In-sample pricing performance is clearly better by recognizing that liquidity effects are present in the pricing of options. This is also the case when we consider all calls together. The evidence is slightly less clear for ITM calls and particularly for 1998. In any case, only in 1998 for ITM calls, the model without liquidity performs better than the model with the liquidity effect (the case of the price below the bid quote). Liquidity, as proxied by the relative bid-ask spread, seems to be an important variable in pricing call options in the Spanish market.

In order to validate our previous results we have carried out a slightly different in-sample test. We separate the sample of each quarter in two subsamples. The first subsample contains 10% of calls for each moneyness degree (*subsample 10*). The other subsample (*subsample 90*) contains the rest of the sample. We estimate the price of each call in the *subsample 10* with the information contained in the *subsample 90*. The bandwidths used are the ones selected by the plug-in method with all available calls in the quarter. This is not the usual in-sample pricing because the calls in the *subsample 10* are not included in their estimation. As in the previous case, in Table 5 we now report both the absolute pricing error and the relative pricing error as a measure of pricing performance. In the first two panels of Table 5 we observe that in all cases both measures are lower when liquidity is incorporated in the estimation of the nonparametric volatility function. We again analyze the proportion of theoretical prices lying outside their corresponding bid-ask spread boundaries. Panel C presents the results when we consider all calls together. Again, the option pricing model with liquidity is superior to the model without liquidity, although in some cases we are not able to reject the null hypothesis. The same holds when we separate the subsample according to moneyness degree¹⁶.

¹⁶ Results are available upon request.

5.2 In-sample pricing performance with alternative smoothing parameters

As mentioned in Section 2.2, we check the appropriateness of the plug-in smoothing parameter estimators by calculating three different bandwidths obtained from the three alternative multivariate criteria described in Appendix C. They are given by expressions (C.4), (C.5) and (C.6).

It should be recognized that our nonparametric problem is inherently a multivariate estimation problem. These alternative criteria -the GCV and the Rice's estimators- have been used in univariate contexts. However, in practice and to the best of our knowledge, they have not been extended to the multivariate case until this paper. They are easier generalized to the multivariate case than the plug-in method.

Table 6 reports the bandwidths obtained for each criteria and for each year during the sample period. In 1996 and 1997, for time to expiration and the spread variable, the bandwidths selected by the multivariate criteria tend to be higher than optimal parameter from the plug-in method. As discussed in Section 2.2, this result may be expected. However, this is not the case for 1998, where the smoothing parameter selected for the moneyness degree is lower than the selected with the plug-in method.

Table 7 compares the effects of the GCV selection method with the plug-in method on option pricing. As we can observe from the table, when we take out the 10% of the sample, we are not able to statistically reject the equality of proportions between the two selection models. This result suggests that the plug-in method produces a quite good estimation of the smoothing parameter. In any case, although not reported, the results in terms of lower proportions of prices lying outside the bid-ask spread boundaries in the model with liquidity remain the same. Of course, further research specifically directed toward this issue is clearly justified¹⁷.

¹⁷ Note that Table 7 is different from Table 4. This is because in Table 4 we price quarterly, with the bandwidths selected for each quarter, and in Table 7 we price with the bandwidths selected yearly.

5.3 Stability of the Risk-Neutral Density and the Out-of-sample Pricing Performance

We next use data and the estimated nonparametric volatility function from one quarter to price options in the following quarter. To justify this out-of-sample analysis is reasonable to argue that a previous stability tests on the risk-neutral densities should be performed. We propose a new stability test to analyze the stability of the risk-neutral densities over all quarters in our sample period.

Aït-Sahalia and Lo (1998a) deal with a similar problem and they derive a diagnostic test based on the integrated squared difference between two risk-neutral densities estimated over two different time periods.

Of course, testing the equality of two nonparametric functions is not a trivial issue. First of all, we work with asymptotic distributions and, secondly, there are several quantities that must be estimated and may affect the results substantially. Practical reasons have led us to propose a simple and easy to implement test based on the so called *randomization tests* to detect if the differences between the risk-neutral densities are statistically significant. This test lies on the permutation of options in our dataset.

More specifically, we assume that options may be randomly assigned to different testing periods. Data are permuted repeatedly and the test statistic shown below is computed for each of the resulting data permutations. These data permutations, including the one representing the obtained results, constitute the reference set for determining the level of significance. The proportion of data permutations in the reference set that have a value of the test stastistic greater than or equal to the value of the experimentally obtained results is the p-value. It should be noted that, in the proposed test, the basis for permuting the data is random assignment. This is why is known as a randomization test¹⁸. The null hypothesis is that the risk-neutral density is the same for each pair of quarters.

The steps follow to implement this tests are as follows:

(i) We define a *grid* of points (returns) where the risk-neutral density is estimated. Note that this is necessary given that the two groups (two periods) do not necessarily have in common the same returns (same set of exercise prices) of the underlying asset. The number of points in the grid is denoted by N_g .

(ii) We estimate the risk-neutral densities, $f_q^*(\cdot), f_{q+1}^*(\cdot)$, using the method described in Section 4, where q denotes a particular given quarter, and liquidity is included.

(iii) We define the test statistic as:

$$T_{\text{obs}} = \sum_{j=1}^{N_g} \left| f_q^*(S_j) - f_{q+1}^*(S_j) \right| \quad (22)$$

(iv) We permute the rows of the data matrix -in our case the set of call options for both quarters q and $q+1$ a given number of times. In this case we employ 200 permutations. In general, we denote B the number of permutations.

(v) For each of the $b = 1, \dots, B$ permutations of the data, we assign N_q options to quarter q and the rest to the following quarter. This constitutes our new dataset originated from permuting the data.

(vi) We estimate $f_q^{*b}(\cdot), f_{q+1}^{*b}(\cdot)$, by keeping the values for the bandwidths we obtained under the original dataset. This seems to be reasonable since the optimal bandwidth is not affected by the randomization process, and computational costs are cheaper.

(vii) We compute the test statistic of equation (22) for each of the $b = 1, \dots, B$ permutations of data.

(viii) We next calculate the p-value or the critical value for our randomization test. The proportion of data permutations in the reference set that has values of the statistic (22) greater than or equal to the value for the experimentally obtained results is the p-value

¹⁸ See Good (1994) for details on randomization tests.

we are looking for. The critical value is the $(1-\alpha)$ th percentile of the values taken by the test statistic calculated from the respective permutations.

(ix) We conclude whether or not the differences in the risk-neutral densities between two consecutive quarters are statistically significant at the α level of significance.

The results are contained in Table 8. They show a remarkable stability between risk-neutral densities for any two consecutive quarters. It should be pointed out that this result is also consistent with the findings of Aït-Sahalia and Lo (1998a), where they are not able to reject the stability for two semesters during the same year. In our data, and for all cases, the null hypothesis of no significant differences between quarters can not be rejected. It may be noticed that in every turn of the year, the p-value obtained is lower than in other cases, although we can not reject the null hypothesis.

This striking result seems to be useful in justifying the use of a given risk-neutral density estimated for a given quarter as the basis for pricing options during the following quarter. The out-of-sample performance is investigated next. We use the nonparametric volatility function estimated for a given quarter q as the true volatility function for the following quarter $q+1$.

We again calculate the theoretical option prices of our two competing models using expression (16) by employing the nonparametric volatility function estimated over the previous quarter. As before, the Z-statistic given in equation (21) is used.

The results are contained in Table 9. Unfortunately, the empirical performance of both models is really disappointing. Independently of the moneyness degree considered, we are not able to reject the equality of performance between the two models. Moreover, the proportions of theoretical prices lying outside the bid-ask spread is considerably higher than in the in-sample case. The conclusion seems to be clear. The out-of-sample performance of our semiparametric models is quite poor either with liquidity or without liquidity.

It should be noticed that the results of Table 9 are based on the comparison between the second and third quarter of each year during the sample period. We report this specific

out-of-sample quarter performance as a representative case of what is found in all quarters. In fact, as a way of aggregating quarter Z statistics over a single year, the quarter Z 's are added and divided by the squared root of the number of quarters where we do the pricing ($\sqrt{4}$ in 1997, $\sqrt{3}$ in 1996 and 1998), to obtain an aggregate $N(0,1)$ statistic from which an aggregate p-value is obtained. These aggregate p-values for differences in proportions between the two competing models are reported in Table 10. In this case, we just include the results for all calls. As it can be observed, we are never able to reject the equality of proportions between the model with liquidity and the model without liquidity. The out-of-sample performance is always poor independently of the model.

This is an interesting result. In spite of the fact that we are not able to reject the stability of the risk-neutral densities over time, the pricing results are very disappointing. This should be understood as a warning about the stability test as a sufficient condition for pricing.

6. Conclusions

This paper has investigated the effects of liquidity, as proxied by the relative bid-ask spread, on the pricing of options. Given the evidence contained in Peña, Rubio and Serna (1999a) where linear (and non-linear) causality tests between the shape of the volatility smile and the bid-ask spread show a bidirectional Granger causality, this paper estimates a nonparametric volatility function in which liquidity is incorporated as a key explanatory variable. Given the structure of our dataset across the degree of moneyness, the paper employs the Symmetrized Nearest Neighbors (SNN) nonparametric estimator rather than the traditional kernel estimator. Moreover, special care is taken in the estimation of the smoothing parameter. The results show that the in-sample performance is clearly favorable to a semiparametric model with liquidity relative to a similar model estimated without liquidity. In fact, liquidity is very useful if we employ options close together on time. This is supported by the in-sample pricing when we take off 10% of data.

Unfortunately, however, the out-of-sample performance results are quite disappointing. We employ the nonparametric volatility function estimated for a given quarter q as the true volatility function for the following quarter $q+1$. Our two semiparametric competing option pricing models present high proportions of theoretical prices lying outside the bid-ask spread boundaries. At the same, we are not able to reject the equality of these proportions across both models. This result is found despite the fact that we do not reject the stability of risk-neutral densities over the quarters covered by the research.

Future work will concentrate in extending the univariate plug-in criterion to a multivariate framework. We will also analyze whether the optimal multivariate plug-in selection method is able to obtain a better trade-off between bias and variance in the estimation of the volatility function around the at-the-money options.

Finally, given our in-sample result, we are planning to analyze the pricing performance of our semiparametric option pricing model with liquidity on daily basis. This is to say, we will daily estimate the nonparametric volatility function with the SNN estimator, and this function will be used as an input in the option pricing function for the following day. This procedure is closer in spirit to other empirical related papers in option pricing.

APPENDIX A
SYMMETRIZED NEAREST NEIGHBORS (SNN) METHODOLOGY

For simplicity, we make a comparison between kernel and SNN estimators in the univariate context. Let us assume the following data generating process,

$$Y_i = m(X_i) + \varepsilon_i ; i = 1, \dots, n \quad (\text{A.1})$$

where $m(\cdot)$ is the unknown function, and X_i are i.i.d. random variables having a density function $f(x)$. The disturbance terms are also assumed to be i.i.d., and a sample of size n is taken. The main advantage of the nonparametric estimation is that we do not have to assume anything about the functional form of the regression function. We only assume that this function is smooth and (at least) twice differentiable. The kernel estimator is:

$$\hat{m}_K(x) = \frac{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) Y_i}{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)} \quad (\text{A.2})$$

where h is the smoothing or bandwidth parameter to be selected.

The kernel or weight function K has the following properties:

$$\begin{aligned} \int K(u) du &= 1 \\ \int K(u) u du &= 0 \\ \int K(u) u^2 du &= d_k \quad (\text{finite}) \\ \int K(u)^2 du &= c_k \quad (\text{finite}) \end{aligned}$$

In practice, we employ the Gaussian kernel which has these properties and is given by,

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

The estimation proposed by Yang (1984) and studied by Stute (1984) is:

$$\hat{m}_S(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{F_n(x) - F_n(X_i)}{h}\right) Y_i \quad (\text{A.3})$$

The known *k-nearest neighbors* estimate define neighbors in terms of the k -nearest points of x . On the other hand, with the SNN methodology, neighbors are defined in terms of the distance between the values of the empirical distribution function at each point. Since expression (A.3) picks up its neighbors symmetrically, it is known as the *Symmetrized Nearest Neighbors* (SNN).

Under the usual assumptions in kernel estimation, it is well known that the leading term in the mean square error is given by¹⁹:

$$\text{MSE}(m_K(x)) = \frac{d_K^2 h^4}{4} \frac{(m''f + 2m'f')^2(x)}{f^2(x)} + \frac{s^2}{nhf(x)} c_K \quad (\text{A.4})$$

where d_K and c_K are constants (already defined above) that depend on the kernel chosen, and s^2 is the variance of the disturbance term.

The MSE for the SNN estimator is:

$$\text{MSE}(m_S(x)) = \frac{d_K^2}{4} h^4 \left(\frac{(m''f - m'f')(x)}{f^3(x)} \right)^2 + \frac{s^2}{nh} c_K \quad (\text{A.5})$$

By minimizing the MSE for each case, the optimal bandwidth h is derived. This is substituted into (A.4) and (A.5) to get the minimum MSE denoted by MSE^* . If we employ the kernel estimator:

$$\text{MSE}^*(\hat{m}_K(x)) = \frac{5}{4} \frac{(d_K^2)^{1/5} (s^2)^{4/5} c_K^{4/5} H(x)^{1/5}}{n^{4/5} f(x)^{4/5}} \quad (\text{A.6})$$

where,

$$H(x) = \frac{(m''f + 2m'f')^2(x)}{f^2(x)}$$

On the other hand, if we use a SNN estimator has an equivalent expression which asymptotically is equal to:

$$\text{MSE}^*(\hat{m}_S(x)) = \frac{5 \left(d_K^2\right)^{1/5} \left(s^2\right)^{4/5} c_K^{4/5} S(x)^{1/5}}{4 n^{4/5}} \quad (\text{A.7})$$

where,

$$S(x) = \frac{(m''f - m'f')^2(x)}{f^6(x)}$$

Thus, we can conclude that it is advisable to employ the SNN estimator instead of the kernel estimator when $\text{MSE}^*(\hat{m}_S(x)) < \text{MSE}^*(\hat{m}_K(x))$. This will be the case for any x such that:

$$\frac{H(x)}{f^4(x)} > S(x) \quad (\text{A.8})$$

Alternatively, condition (A.8) is equivalent to:

$$(m''f + 2m'f')^2 > (m''f - m'f')^2 \Leftrightarrow m'f'(m'f' + 2m''f) > 0 \quad (\text{A.9})$$

As we already mentioned, depending on the specific functional form of the unknown function, $m(x)$, and the density function, $f(x)$, we would prefer either one of them. This is shown in the following simple example. The intuition behind this example is that it seems reasonable to expect that a typical smile volatility pattern (if it exists) may be captured by a (parametric) quadratic regression equation. At the same, it is not plausible to assume that the design of moneyness is fixed during the period of study. Hence, the assumption of a normal density for x seems appropriate.

¹⁹ $\text{MSE} = (\text{Bias})^2 + \text{Variance}$.

Suppose that we have the following parametric model:

$$y_i = m(x_i) + \varepsilon_i$$

where

$$m(x_i) = ax_i^2$$

Let us also assume that the errors are i.i.d. and that x have a normal density with mean 0 and variance s^2 :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/s^2}$$

By taking the appropriate derivatives for $m(x)$ and $f(x)$ and substituting them into expression (A.9), we obtain that the MSE under the kernel estimation is higher than in the SNN estimation as long as $|x| > s$. In words, when the points are in a distance from the mean (in this example zero, but it may be generalized) higher than the standard deviation, we obtain better results using the SNN estimation.

This example suggests that whenever we are interested in estimating in those places where the density is very small (away from the mean), the SNN may yield better results than the traditional kernel estimators. Given the empirical distribution of moneyness, this suggest that, in general, the $MSE^*(\hat{m}_S(x))$ might be lower in those sections of the smile where the degree of moneyness is far away from at-the-money options. Of course, as we said before, whether in general $MSE^*(\hat{m}_S(x)) < (>) MSE^*(\hat{m}_K(x))$ depends on the particular functional form of the unknown function $m(\cdot)$ and the density $f(x)$. We suspect that in most cases the SNN estimator would present better results when estimating in places where the density is small.

APPENDIX B

APPROPRIATE DERIVATIVES FOR THE CALCULATION OF THE RISK-NEUTRAL DENSITY

The Black-Scholes formula for a European futures call option is²⁰:

$$c = e^{-r(T-t)} [FN(d_1) - XN(d_2)]$$

where,

$$d_1 = \frac{\ln(F/X) + \sigma^2/2(T-t)}{\sigma\sqrt{T-t}}$$

$$d_2 = \frac{\ln(F/X) - \sigma^2/2(T-t)}{\sigma\sqrt{T-t}} = d_1 - \sigma\sqrt{T-t}$$

$N(d_1)$ is the value for the accumulative probability distribution of a normal variable with mean 0 and variance 1. Its derivative is, therefore, the normal density function:

$$N'(d_1) = \frac{1}{\sqrt{2\pi}} e^{-d_1^2/2}$$

It is easy to show that,

$$N'(d_2) = \frac{1}{\sqrt{2\pi}} e^{-d_2^2/2} = N'(d_1) \frac{F}{X}$$

Our objective is to find the second derivative of the call price with respect to the exercise price. It should be noted that, in our case, the volatility is a function of the exercise price too:

$$\frac{\partial^2 \hat{c}}{\partial X^2} = \frac{\partial^2 c}{\partial X^2} + 2 \frac{\partial^2 c}{\partial X \partial \sigma} \frac{\partial \hat{c}}{\partial X} + \left[\frac{\partial^2 c}{\partial \sigma^2} + \frac{\partial c}{\partial \sigma} \right] \frac{\partial^2 \hat{\sigma}}{\partial X^2} \quad (\text{B.1})$$

We need the following results,

²⁰ In fact, this is Black (1976).

$$\frac{\partial c}{\partial X} = e^{-r(T-t)} N(d_2)$$

$$\frac{\partial^2 c}{\partial X^2} = e^{-r(T-t)} \frac{1}{X \sqrt{2\pi\sigma^2(T-t)}} \exp\left\{-\frac{[\ln(F/X) - \sigma^2/2(T-t)]^2}{2\sigma^2(T-t)}\right\}$$

$$\frac{\partial d_1}{\partial X} = \frac{-1}{X\sigma\sqrt{T-t}}$$

$$\frac{\partial c}{\partial \sigma} = e^{-r(T-t)} N'(d_1) F \sqrt{T-t}$$

$$\frac{\partial c}{\partial \sigma \partial X} = -e^{-r(T-t)} \frac{F}{X} N''(d_1) \frac{1}{\sigma}$$

$$\frac{\partial^2 c}{\partial \sigma^2} = -e^{-r(T-t)} F \frac{d_2}{\sigma} \sqrt{T-t} N''(d_1)$$

We need to calculate the derivatives of our volatility function estimation with respect to the exercise price, $\partial \hat{\sigma} / \partial X$ and $\partial^2 \hat{\sigma} / \partial X^2$. To obtain them, a kernel estimation is used. But recall that our nonparametric estimation of volatility depends on moneyness, and it does not depend on the exercise price. Hence, the appropriate derivatives for the volatility with respect to the exercise price are:

$$\frac{\partial^m \hat{\sigma}}{\partial X^m} = \frac{\partial^m \hat{\sigma}}{\partial (X/F)^m} \frac{1}{F^m}$$

Therefore,

$$\begin{aligned}
\frac{\partial^2 \hat{c}}{\partial X^2} &= e^{-r(T-t)} \frac{1}{X \sqrt{2\pi\sigma^2(T-t)}} \exp\left\{-\frac{[\ln(F/X) - \sigma^2/2(T-t)]^2}{2\sigma^2(T-t)}\right\} \\
&- 2e^{-r(T-t)} \frac{F}{X} N''(d_1) \frac{1}{\sigma} \frac{\partial \hat{c}}{\partial(X/F)} \frac{1}{X} \\
&+ \left[-e^{-r(T-t)} F \frac{d_2}{\sigma} \sqrt{T-t} N''(d_1) + e^{-r(T-t)} N'(d_1) F \sqrt{T-t} \right] \frac{\partial^2 \hat{c}}{\partial(X/F)^2} \frac{1}{F^2}
\end{aligned} \tag{B.2}$$

APPENDIX C
SELECTING THE SMOOTHING PARAMETER

We know that the accuracy of kernel smoothers as estimators of any function $m(\cdot)$, as in expression (A.1) is a function of the kernel and the bandwidth h . In practice, it is well accepted that the accuracy depends mainly on the chosen smoothing parameter h . In this appendix we first discuss the plug-in method as applied to our multivariate particular case. Secondly, we present the multivariate cross-validation criteria also employed in our estimations.

Let us briefly describe the appropriate algorithm for the simple kernel univariate case. We know that asymptotic mean squared error for the usual kernel estimator is given by the following equation:

$$\text{AMSE} = \frac{1}{nh} s^2 c_K + \frac{h^4 d_K^2 [m''(x)]^2}{4} \quad (\text{C.1})$$

and the asymptotic integrated mean squared error is:

$$\text{AMISE} = \frac{1}{nh} s^2 c_K + h^4 d_K^2 \left[\int m''(x)^2 dx \right] \frac{1}{4} \quad (\text{C.2})$$

$$\frac{\partial \text{AMISE}}{\partial h} = 0 \Rightarrow h = \left(\frac{1}{n} \right)^{1/5} \left(\frac{c_K}{d_K^2} \right)^{1/5} \left(\frac{s^2}{\int m''(x)^2 dx} \right)^{1/5}$$

where the problem is that $m''(x)$ is unknown and must be estimated. Note that this is the case given that the constants c_k and d_k depend on the kernel function assumed, n is the sample size, and the value of the variance of the response variable may easily be estimated as,

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

Let $\hat{m}_2(x, h_2)$ be the estimator of $m''(x)$ where h_2 refers to the second derivative. It turns out that this can be estimated using a kernel estimator again

$$\hat{m}_2(x, h_2) = \frac{1}{nh_2^3} \sum_{i=1}^n K''\left(\frac{x - X_i}{h_2}\right) Y_i \quad (\text{C.3})$$

where the specific kernel assumed must have the following properties:

$$\begin{aligned} \int K''(u) du &= 0 \\ \int K''(u) u du &= 0 \\ \int K''(u) u^2 du &= 2 \\ \int K''(u) u^3 du &= 0 \\ \int K''(u) u^4 du &= d_k \text{ (finite)} \\ \int K''(u)^2 du &= c_k \text{ (finite)} \end{aligned}$$

In particular, the kernel function $K''(u) = \frac{105}{16}(-5u^4 + 6u^2 - 1)$ has these properties and has been used in this estimation.

The bandwidth h_2 is estimated directly according to the algorithm proposed by Gasser, Kneip, and Köhler (1991) for the univariate case. The algorithm is performed by the following steps:

$$\begin{aligned} \text{(i)} \quad \hat{h}_0 &= \frac{1}{n} \\ \text{(ii)} \quad \hat{h}_j &= \left(\frac{c_k \hat{s}^2}{nd_k^2 \int \hat{m}_2(x, \hat{h}_{j-1} n^{1/10})^2 dx} \right)^{1/5} \\ \text{(iii)} \quad &\text{stop when it converges} \end{aligned}$$

In words,

- Obtain the expression for the mean integrated squared error for the nonparametric regression (expression C.2)
- Estimate the partial derivatives involved using pilot bandwidths. In this case we need the second derivative of the $m(x)$ function, where the pilot bandwidth is h_2 .
- Get the initial bandwidth and transform it to obtain another pilot bandwidths.
- Iterate until convergence.

The extension of this univariate plug-in method to the multivariate case is not trivial. For this reason, in order to check the robustness of our results, we employ the (easier) extensions of the univariate cross-validation criteria given in Härdle (1990). In particular, the criteria employed in this research and their penalizing functions are the following:

(i) The Generalized Cross-Validation (GCV), where we have to minimize:

$$\text{GCV}(h_1, h_2, h_3) = \text{RSS}(h_1, h_2, h_3) \frac{1}{\left[1 - \frac{1}{n} \frac{K(0)^3}{h_1 h_2 h_3}\right]^2} \quad (\text{C.4})$$

(ii) Rice's bandwidth selector I, where we have to minimize:

$$\text{RICE}_I(h_1, h_2, h_3) = \text{RSS}(h_1, h_2, h_3) \frac{1}{1 - \frac{2}{n} \frac{K(0)^3}{h_1 h_2 h_3}} \quad (\text{C.5})$$

(iii) Rice's bandwidth selector II, where we have to minimize:

$$\text{RICE}_{II}(h_1, h_2, h_3) = \text{RSS}(h_1, h_2, h_3) - \frac{2\hat{s}^2}{n} \frac{K(0)^3}{h_1 h_2 h_3} \quad (\text{C.6})$$

where $\text{RSS}(\cdot)$ is the sum of squared residuals; this is to say, the residuals we obtain in the nonparametric estimation with the kernel. Finally, $K(0)$ is the kernel evaluated at 0.

It is important to point out that either (C.4), (C.5) or (C.6) are truly multivariate bandwidth selection methods.

REFERENCES

- Aït-Sahalia, Y., and A. Lo (1998a). "Nonparametric estimation of state-price densities implicit in financial asset prices", *Journal of Finance* 53, pp.499-547.
- Aït-Sahalia, Y., and A. Lo (1998b). "Nonparametric risk management and implied risk aversion", Working Paper, Graduate School of Business, University of Chicago.
- Bakshi, G., Cao, C., and Z. Chen (1997). "Empirical performance of alternative option pricing models", *Journal of Finance* 52, pp. 2003-2049.
- Bates, D. (1996). "Jumps and stochastic volatility: exchange rate processes implicit in Deutsche mark options", *Review of Financial Studies* 9, pp. 69-107.
- Black, F. (1976). "The pricing of commodity contracts", *Journal of Financial Economics* 3, pp. 167-179.
- Black, F. and M. Scholes (1973). "The pricing of options and corporate liabilities", *Journal of Political Economy* 81, pp. 637-659.
- Chernov, M. and E. Ghysels (1998). "What data should be used to price options?", unpublished manuscript, Pennsylvania State University.
- Das, S., and R. Sundaram (1998). "Of smiles and smirks: A term-structure perspective", forthcoming in the *Journal of Financial and Quantitative Analysis*.
- Dumas, B., J. Fleming and R. Whaley (1998). "Implied volatility functions: empirical tests", *Journal of Finance*, 53, pp. 2059-2106.
- Fiorentini, G., León, A. and G. Rubio (1998). "Short-term options with stochastic volatility: estimation and empirical performance", Working Paper, Instituto Valenciano de Investigaciones Económicas, Spain.

Gasser, T., Kneip, A. and W. Köhler (1991). "A flexible and fast method for automatic smoothing", *Journal of the American Statistical Association*, 86, pp. 643-652.

Glosten, L., Jagannathan, R. and D. Runkle (1993). "On the relation between the expected value and the volatility of the nominal excess return on stocks", *Journal of Finance* 48, pp. 1779-1801.

Good, P. (1994). "Permutation Tests. A practical guide to resampling methods for testing hypotheses", Springer-Verlag, New York.

Härdle, W. (1990). "Applied nonparametric regression", Cambridge University Press, Cambridge.

Heston, S. (1993). "A closed-form solution for options with stochastic volatility with applications to bond and currency options", *Review of Financial Studies* 6, pp. 327-344.

León, A. and G. Rubio (1999). "Smiling under stochastic volatility", Working Paper, University of Alicante.

Longstaff, F. (1995). "Option pricing and the martingale restriction", *Review of Financial Studies* 8, pp. 1091-1124.

Peña, I., Rubio, G. and G. Serna (1999a). "Why do we smile? On the determinants of the implied volatility function", *Journal of Banking and Finance* 23, pp. 1151-1179.

Peña, I., Rubio, G. and G. Serna (1999b). "Smiles and transaction costs in option pricing", Working Paper, Universidad Carlos III.

Rubinstein, M. (1994). "Implied binomial trees", *Journal of Finance* 49, pp. 771-818.

Shreve, S., Chalasani, P. and S. Jha (1997). "Stochastic calculus and finance", Unpublished manuscript, Carnegie Mellon University.

Stute, W. (1984). "Asymptotic normality of nearest neighbor regression function estimates", *Annals of Statistics* 12, pp. 917-926.

Taylor, S. J. and X. Xu (1994). "The magnitude of implied volatility smiles: theory and empirical evidence for exchange rates", *The Review of Futures Markets* 13, pp. 355-380.

Yang, S. (1981). "Linear functions of concomitants of order statistics with applications to nonparametric estimation of a regression function", *Journal of the American Statistical Association* 76, pp. 658-662.

TABLE 1**SAMPLE CHARACTERISTICS OF IBEX-35 FUTURES OPTIONS**

Average prices, average relative bid-ask spread and the number of available calls are reported for each moneyness category. All call options transacted over the 45 minute interval from 16:00 to 16:45 are employed from January 2, 1996 to November 10, 1998. X is the exercise price and F denotes the futures price of the IBEX-35 index. Moneyness is defined as the ratio of the exercise price to the futures price. OTM, ATM, and ITM are out-of-the-money, at-the-money, and in-the-money options respectively.

CALLS				
	Moneyness (K/F)	Average Price	Average Bid-Ask Spread	Number of Observations
DEEP OTM:	1.03-1.08	61.6	0.378	1478
OTM:	1.01-1.03	81.7	0.237	1457
ATM:	0.99-1.01	120.4	0.174	1422
ITM:	0.97-0.99	180.0	0.122	355
DEEP ITM:	0.90-0.97	381.2	0.124	86
ALL CALLS:	-	99.6	0.251	4798
PUTS				
	Moneyness (K/F)	Average Price	Average Bid-Ask Spread	Number of Observations
DEEP OTM:	1.03-1.08	461.7	0.197	53
OTM:	1.01-1.03	165.2	0.119	180
ATM:	0.99-1.01	126.5	0.152	878
ITM:	0.97-0.99	90.6	0.196	988
DEEP ITM:	0.90-0.97	64.1	0.325	1424
ALL PUTS:	-	98.2	0.233	3523

TABLE 2**SMOOTHING PARAMETERS CALCULATED BY THE ITERATIVE PLUG-IN METHOD:
THE NONPARAMETRIC VOLATILITY FUNCTION**

All call option data for each year separately is used in calculating the optimal bandwidth parameters for each of our three explanatory variables when estimating the nonparametric volatility function, $\hat{\sigma}(\xi, SP, \tau)$. We follow the univariate iterative procedure suggested by Gasser, Kneip, and Köhler (1991) in which the optimal bandwidth parameter, h_j , ($j = \xi, SP, \tau$) is obtained by minimizing the mean integrated squared error.

BANDWIDTH PARAMETERS	1996	1997	1998
h_ξ (moneyness)	0.0537	0.0614	0.0931
h_{SP} (bid-ask spread)	0.0383	0.0434	0.0721
h_τ (time to expiration)	0.0397	0.0502	0.0941

TABLE 3

IN-SAMPLE ABSOLUTE AND PERCENTAGE PRICING ERRORS FOR ALTERNATIVE SEMIPARAMETRIC OPTION PRICING MODELS: THE LIQUIDITY EFFECTS

The nonparametric volatility function is estimated for each quarter from January 2, 1996 to November 10, 1998 using all available call options in each quarter. The corresponding call price is calculated by using the Black-Scholes pricing function evaluated at the previously (nonparametrically) estimated volatility. The bandwidths parameters are obtained by the plug-in method. The reported absolute pricing error is the sample average of the squared difference between the model price and the market price for each call in a given moneyness category. The reported percentage pricing error is the sample average of the squared difference between the model price and the market price divided by the market price for each call in a given moneyness category. We compare the pricing errors between the semiparametric option pricing model with liquidity (WL) and the semiparametric option pricing model without liquidity (WOL). For presentation reasons, the empirical results are aggregated for years. Moneyness is defined as the ratio of the exercise price to the futures price. OTM, ATM, and ITM are out-of-the-money, at-the-money, and in-the-money options respectively.

PANEL A: ABSOLUTE PRICING ERRORS FOR CALLS						
	1996		1997		1998	
	WL	WOL	WL	WOL	WL	WOL
DEEP OTM	0.523	1.985	3.909	8.713	24.816	41.548
OTM	1.351	2.671	3.919	9.362	20.026	32.404
ATM	1.138	2.932	5.343	11.866	22.823	38.489
ITM	0.711	2.803	4.065	11.487	22.499	27.186
DEEP ITM	0.444	2.095	4.592	15.810	3.324	6.492
ALL CALLS	1.123	2.724	4.031	10.096	22.730	37.420
PANEL B: PERCENTAGE PRICING ERRORS FOR CALLS						
	1996		1997		1998	
	WL	WOL	WL	WOL	WL	WOL
DEEP OTM	0.072	0.221	0.083	0.210	2.082	2.312
OTM	0.206	0.393	0.068	0.149	0.148	0.217
ATM	0.026	0.061	0.038	0.087	0.082	0.151
ITM	0.006	0.031	0.019	0.061	0.055	0.070
DEEP ITM	0.002	0.012	0.015	0.038	0.006	0.013
ALL CALLS	0.012	0.247	0.066	0.159	1.333	1.485

TABLE 4

IN-SAMPLE STATISTICAL SIGNIFICANCE OF PERFORMANCE FOR ALTERNATIVE SEMIPARAMETRIC OPTION PRICING MODELS: THE LIQUIDITY EFFECTS

The nonparametric volatility function is estimated for each quarter from January 2, 1996 to November 10, 1998 using all available call options in each quarter. The corresponding call price is calculated by using the Black-Scholes pricing function evaluated at the previously (nonparametrically) estimated volatility. The bandwidths parameters are obtained by the plug-in method. The statistical performance for pricing errors is assessed by analyzing the proportion of theoretical prices lying outside their corresponding bid-ask spread boundaries. The Z-statistic for testing the differences between two proportions is employed. The statistic is asymptotically distributed as a standardized normal variable. We report the statistical significance of pricing errors between the semiparametric option pricing model with liquidity (WL) and the semiparametric option pricing model without liquidity (WOL). For presentation reasons, the empirical results are aggregated for years. Moneyness is defined as the ratio of the exercise price to the futures price. OTM, ATM, and ITM are out-of-the-money, at-the-money, and in-the-money options respectively.

CATEGORIES	1996			1997			1998		
	WL	WOL	Z-STAT	WL	WOL	Z-STAT	WL	WOL	Z-STAT
<u>OTM CALLS:</u>									
Bid > c > Ask	0.110	0.216	-6.89 (0.000)	0.171	0.417	-17.07 (0.000)	0.306	0.430	-7.39 (0.000)
c > Ask	0.064	0.135	-5.56 (0.000)	0.086	0.237	-12.13 (0.000)	0.156	0.232	-5.30 (0.000)
c < Bid	0.046	0.081	-3.42 (0.000)	0.084	0.179	-8.45 (0.000)	0.149	0.197	-3.57 (0.000)
<u>ATM CALLS:</u>									
Bid > c > Ask	0.100	0.217	-7.18 (0.000)	0.180	0.357	-7.71 (0.000)	0.245	0.411	-6.03 (0.000)
c > Ask	0.050	0.115	-5.19 (0.000)	0.117	0.221	-5.20 (0.000)	0.141	0.220	-3.38 (0.000)
c < Bid	0.050	0.101	-4.31 (0.000)	0.101	0.224	-4.48 (0.000)	0.103	0.191	-3.98 (0.000)
<u>ITM CALLS:</u>									
Bid > c > Ask	0.139	0.213	-3.44 (0.000)	0.137	0.391	-6.10 (0.000)	0.302	0.372	-1.33 (0.091)
c > Ask	0.054	0.114	-2.65 (0.004)	0.036	0.166	-4.11 (0.000)	0.151	0.232	-1.78 (0.037)
c < Bid	0.084	0.129	-1.89 (0.029)	0.101	0.224	-3.46 (0.000)	0.151	0.139	0.311 (0.622)
<u>ALL CALLS:</u>									
Bid > c > Ask	0.109	0.220	-10.51 (0.000)	0.170	0.400	-19.56 (0.000)	0.290	0.421	-9.47 (0.000)
c > Ask	0.057	0.124	-8.05 (0.000)	0.090	0.228	-13.70 (0.000)	0.152	0.229	-6.53 (0.000)
c < Bid	0.052	0.095	-5.77 (0.000)	0.080	0.172	-10.15 (0.000)	0.138	0.192	-4.90 (0.000)

TABLE 5
THE 10% IN-SAMPLE PRICING ERRORS FOR ALTERNATIVE SEMIPARAMETRIC
OPTION PRICING MODELS: THE LIQUIDITY EFFECTS

The nonparametric volatility function is estimated from January 2, 1996 to November 10. The price of each call is estimated in a subsample containing 10% of all available calls. They are estimated using the remaining 90% of calls. The corresponding call price is calculated by using the Black-Scholes pricing function evaluated at the previously (nonparametrically) estimated volatility. The bandwidths parameters are obtained by the plug-in method. Panel A: The reported absolute pricing error is the sample average of the squared difference between the model price and the market price for each call in a given moneyness category. Panel B: The reported percentage pricing error is the sample average of the squared difference between the model price and the market price divided by the market price for each call in a given moneyness category. Panel C: The statistical performance for pricing errors is assessed by analyzing the proportion of theoretical prices lying outside their corresponding bid-ask spread boundaries. The Z-statistic for testing the differences between two proportions is employed. The statistic is asymptotically distributed as a standardized normal variable. The empirical results are aggregated for years. Moneyness is defined as the ratio of the exercise price to the futures price. OTM, ATM, and ITM are out-of-the-money, at-the-money, and in-the-money options respectively. We compare the pricing errors between the semiparametric option pricing model with liquidity (WL) and the semiparametric option pricing model without liquidity (WOL).

PANEL A: ABSOLUTE PRICING ERRORS FOR CALLS

	1996		1997		1998	
	WL	WOL	WL	WOL	WL	WOL
DEEP OTM	0.689	2.029	11.880	15.186	39.265	47.205
OTM	2.037	3.221	13.789	18.284	32.575	37.702
ATM	3.268	5.129	13.459	19.394	19.428	33.007
ITM	1.594	2.438	3.082	10.143	5.591	16.596
DEEP ITM	0.016	0.143	15.086	21.028	1.530	3.941
ALL CALLS	2.521	3.997	12.594	17.095	32.144	40.135

PANEL B: PERCENTAGE PRICING ERRORS FOR CALLS

	1996		1997		1998	
	WL	WOL	WL	WOL	WL	WOL
DEEP OTM	0.102	0.133	0.233	0.288	1.228	1.568
OTM	0.113	0.149	0.164	0.231	0.303	0.323
ATM	0.068	0.107	0.116	0.149	0.106	0.164
ITM	0.016	0.026	0.018	0.051	0.018	0.051
DEEP ITM	0.000	0.000	0.037	0.052	0.004	0.010
ALL CALLS	0.088	0.121	0.179	0.231	0.807	1.026

PANEL C: PERFORMANCE FOR ALTERNATIVE MODELS

CATEGORIES	WL	WOL	Z-STAT	WL	WOL	Z-STAT	WL	WOL	Z-STAT
<u>ALL CALLS:</u>									
Bid > c > Ask	0.196	0.344	-3.80 (0.000)	0.512	0.668	-4.28 (0.000)	0.438	0.570	-2.93 (0.001)
c > Ask	0.108	0.175	-2.16 (0.015)	0.295	0.403	-2.84 (0.002)	0.272	0.338	-1.53 (0.063)
c < Bid	0.087	0.169	-2.63 (0.004)	0.216	0.265	-1.40 (0.080)	0.165	0.231	-1.72 (0.043)

TABLE 6

A COMPARISON OF ALTERNATIVE SMOOTHING PARAMETERS:
THE NONPARAMETRIC VOLATILITY FUNCTION

All call option data for each year separately is used in calculating the optimal bandwidth parameters for each of our three explanatory variables when estimating the nonparametric volatility function, $\hat{\sigma}(\xi, SP, \tau)$. We compare the univariate iterative procedure suggested by Gasser, Kneip, and Köhler (1991) in which the optimal bandwidth parameter is obtained by minimizing the mean integrated squared error, the multivariate the Generalized Cross-Validation (GCV) in which the RSS is penalized by:

$1/\left[1 - 1/n\left(K(0)^3/h_\xi h_{SP} h_\tau\right)\right]^2$, the multivariate Rice's bandwidth selector (I), with penalty equal to $1/1 - 2/n\left(K(0)^3/h_\xi h_{SP} h_\tau\right)$, and the multivariate Rice's bandwidth selector (II), in which the RSS is penalized by $-2\hat{s}^2/n\left(K(0)^3/h_\xi h_{SP} h_\tau\right)$ where s^2 is the variance of the residuals of the kernel regression.

1996				
BANDWIDTH PARAMETERS	PLUG-IN	GCV	RICE I	RICE II
h_ξ (moneyness)	0.0537	0.0537	0.0537	0.0537
h_{SP} (bid-ask spread)	0.0383	0.0882	0.0882	0.0729
h_τ (time to expiration)	0.0397	0.0915	0.0915	0.0756
1997				
BANDWIDTH PARAMETERS	PLUG-IN	GCV	RICE I	RICE II
h_ξ (moneyness)	0.0614	0.0614	0.1166	0.0614
h_{SP} (bid-ask spread)	0.0434	0.0998	0.0824	0.0998
h_τ (time to expiration)	0.0502	0.0502	0.0502	0.0502
1998				
BANDWIDTH PARAMETERS	PLUG-IN	GCV	RICE I	RICE II
h_ξ (moneyness)	0.0931	0.0465	0.0838	0.0465
h_{SP} (bid-ask spread)	0.0721	0.0721	0.0721	0.0721
h_τ (time to expiration)	0.0941	0.0941	0.0847	0.0941

TABLE 7

THE 10% IN-SAMPLE PRICING PERFORMANCE FOR THE SEMIPARAMETRIC OPTION PRICING MODEL WITH LIQUIDITY FOR ALTERNATIVE SMOOTHING PARAMETERS:

The nonparametric volatility function is estimated yearly during the sample period. The price of each call is estimated in a subsample containing 10% of all available calls. They are estimated using the remaining 90% of calls. The corresponding call price is calculated by using the Black-Scholes pricing function evaluated at the previously (nonparametrically) estimated volatility. We compare the statistical significance of the in-sample pricing performance when the following smoothing parameter selection procedures are employed: the univariate iterative procedure suggested by Gasser, Kneip, and Köhler (1991) in which the optimal bandwidth parameter is obtained by minimizing the mean integrated squared error and the multivariate the Generalized Cross-Validation (GCV) in which the RSS is penalized by: $1/\left[1 - 1/n\left(K(0)^3/h_\xi h_{SP} h_\tau\right)\right]^2$. The statistical performance is assessed by analyzing the proportion of theoretical prices lying outside their corresponding bid-ask spread boundaries. The Z-statistic for testing the differences between two proportions is employed. The statistic is asymptotically distributed as a standardized normal variable.

	1996			1997			1998		
	Plug-in	GCV	Z-STAT	Plug-in	GCV	Z-STAT	Plug-in	GCV	Z-STAT
<u>ALL CALLS:</u>									
Bid > c > Ask	0.444	0.549	-2.59 (0.004)	0.697	0.732	-1.03 (0.151)	0.743	0.743	0.00 (0.500)
c > Ask	0.268	0.326	-1.55 (0.006)	0.383	0.413	-0.77 (0.220)	0.493	0.493	0.00 (0.500)
c < Bid	0.176	0.222	-1.36 (0.086)	0.314	0.319	-0.19 (0.436)	0.250	0.250	0.00 (0.500)

TABLE 8**A RANDOMIZATION TEST FOR THE STABILITY OF DENSITIES OVER TIME**

A new test to analyze the stability of the risk-neutral density over different quarters is proposed. We basically assume that options may be randomly assigned to different quarters. Data are permuted repeatedly and the test statistic shown below is computed for each of the resulting data permutations. These data permutations, including the one representing the obtained results, constitute the reference set for determining the level of significance. The proportion of data permutations in the reference set that have a value of the test statistic greater than or equal to the value of the experimentally obtained results is the p-value. It should be noted that, in the proposed test, the basis for permuting the data is random assignment. This is why is known as a randomization test. The null hypothesis is that the risk-neutral density is the same for each pair of quarters. The test statistic is defined as:

$$T_{\text{obs}} = \sum_{j=1}^{N_g} \left| f_q^*(S_j) - f_{q+1}^*(S_j) \right|$$

where N_g is the number of time points in the previously defined grid where the risk-neutral densities are estimated, and q refers to quarter.

QUARTERS	VALUE OF THE TEST STATISTIC	P-VALUE	VALUE AT 10%
1st 96 = 2nd 96	0.3123	0.790	0.793
2nd 96 = 3rd 96	0.3748	0.730	0.783
3rd 96 = 4th 96	0.2818	0.510	0.496
4th 96 = 1st 97	0.5101	0.130	0.562
1st 97 = 2nd 97	0.2154	0.735	0.452
2nd 97 = 3rd 97	0.2635	0.655	0.516
3rd 97 = 4th 97	0.4031	0.430	0.680
4th 97 = 1st 98	0.1972	0.275	0.250
1st 98 = 2nd 98	0.1413	0.770	0.353
2nd 98 = 3rd 98	0.2388	0.715	0.449

TABLE 9
OUT-OF-SAMPLE STATISTICAL SIGNIFICANCE OF PERFORMANCE FOR
ALTERNATIVE SEMIPARAMETRIC OPTION PRICING MODELS: THE LIQUIDITY
EFFECTS

The nonparametric volatility function is estimated for each quarter from January 2, 1996 to November 10, 1998 using all available call options in each quarter. The corresponding call price is calculated by using the Black-Scholes pricing function evaluated at the previously (nonparametrically) estimated volatility. The bandwidths parameters are obtained by the plug-in method. The statistical performance for pricing errors is assessed by analyzing the proportion of theoretical prices lying outside their corresponding bid-ask spread boundaries. The Z-statistic for testing the differences between two proportions in each quarter is employed. The statistic is asymptotically distributed as a standardized normal variable. We report the statistical significance of pricing errors between the semiparametric option pricing model with liquidity (WL) and the semiparametric option pricing model without liquidity (WOL), using data and the estimated nonparametric volatility function from one quarter to price options in the following quarter. For presentation reasons, the empirical results are reported only for the 2nd vs. the 3rd. quarter of each year. Moneyness is defined as the ratio of the exercise price to the futures price. OTM, ATM, and ITM are out-of-the-money, at-the-money, and in-the-money options respectively.

CATEGORIES	1996 (2nd vs. 3rd)			1997 (2nd vs. 3rd)			1998 (2nd vs. 3rd)		
	WL	WOL	Z-STAT	WL	WOL	Z-STAT	WL	WOL	Z-STAT
<u>OTM CALLS:</u>									
Bid > c > Ask	0.670	0.592	2.12 (0.982)	0.789	0.778	0.50 (0.691)	0.915	0.907	0.54 (0.705)
c > Ask	0.111	0.139	-1.07 (0.142)	0.282	0.277	0.23 (0.591)	0.384	0.382	0.10 (0.539)
c < Bid	0.558	0.452	2.85 (0.997)	0.506	0.501	0.21 (0.583)	0.530	0.525	0.21 (0.583)
<u>ATM CALLS:</u>									
Bid > c > Ask	0.569	0.512	1.26 (0.896)	0.771	0.638	2.84 (0.997)	0.838	0.838	0.00 (0.500)
c > Ask	0.227	0.138	2.87 (0.997)	0.180	0.085	3.48 (0.999)	0.462	0.451	0.20 (0.582)
c < Bid	0.341	0.373	-0.74 (0.229)	0.590	0.552	0.78 (0.782)	0.376	0.387	-0.21 (0.417)
<u>ITM CALLS:</u>									
Bid > c > Ask	0.703	0.741	-0.43 (0.333)	0.400	0.600	-2.23 (0.012)	0.937	0.937	0.00 (0.500)
c > Ask	0.148	0.148	0.00 (0.500)	0.033	0.200	-2.28 (0.011)	0.937	0.937	0.00 (0.500)
c < Bid	0.555	0.592	-0.39 (0.318)	0.366	0.400	-0.37 (0.355)	0.000	0.000	0.00 (0.500)
<u>ALL CALLS:</u>									
Bid > c > Ask	0.635	0.574	2.23 (0.987)	0.762	0.737	1.22 (0.888)	0.901	0.894	0.45 (0.673)
c > Ask	0.158	0.139	0.95 (0.828)	0.245	0.231	0.74 (0.770)	0.419	0.415	0.18 (0.571)
c < Bid	0.477	0.434	1.55 (0.939)	0.516	0.506	0.44 (0.670)	0.481	0.479	0.09 (0.535)

TABLE 10

**OUT-OF-SAMPLE STATISTICAL SIGNIFICANCE OF PERFORMANCE FOR
ALTERNATIVE SEMIPARAMETRIC OPTION PRICING MODELS:
AGGREGATING P-VALUES FOR THE DIFFERENCES IN PROPORTIONS OF
THEORETICAL PRICES LYING OUTSIDE THE BID-ASK BOUNDARIES**

The nonparametric volatility function is estimated for each quarter from January 2, 1996 to November 10, 1998 using all available call options in each quarter. The corresponding call price is calculated by using the Black-Scholes pricing function evaluated at the previously (nonparametrically) estimated volatility. The bandwidths parameters are obtained by the plug-in method. The statistical performance for pricing errors is assessed by analyzing the proportion of theoretical prices lying outside their corresponding bid-ask spread boundaries. The Z-statistic for testing the differences between two proportions in each quarter is employed. The statistic is asymptotically distributed as a standardized normal variable. We report the statistical significance of pricing errors between the semiparametric option pricing model with liquidity (WL) and the semiparametric option pricing model without liquidity (WOL), using data and the estimated nonparametric volatility function from one quarter to price options in the following quarter. We add the Z-statistic corresponding to each quarter every year, the sum is divided by the square root of 4, and the associated p-value to this aggregate Z-statistic is reported.

	1996	1997	1998
	AGGREGATE P-VALUE (from quarters)	AGGREGATE P-VALUE (from quarters)	AGGREGATE P-VALUES (from quarters)
<u>ALL CALLS:</u>			
Bid > c > Ask	0.994	0.992	0.998
c > Ask	0.963	0.935	0.999
c < Bid	0.976	0.974	0.999

**Figure 1: Nonparametric Volatility Smiles in 1998:
Kernel Estimator versus the SNN estimator**

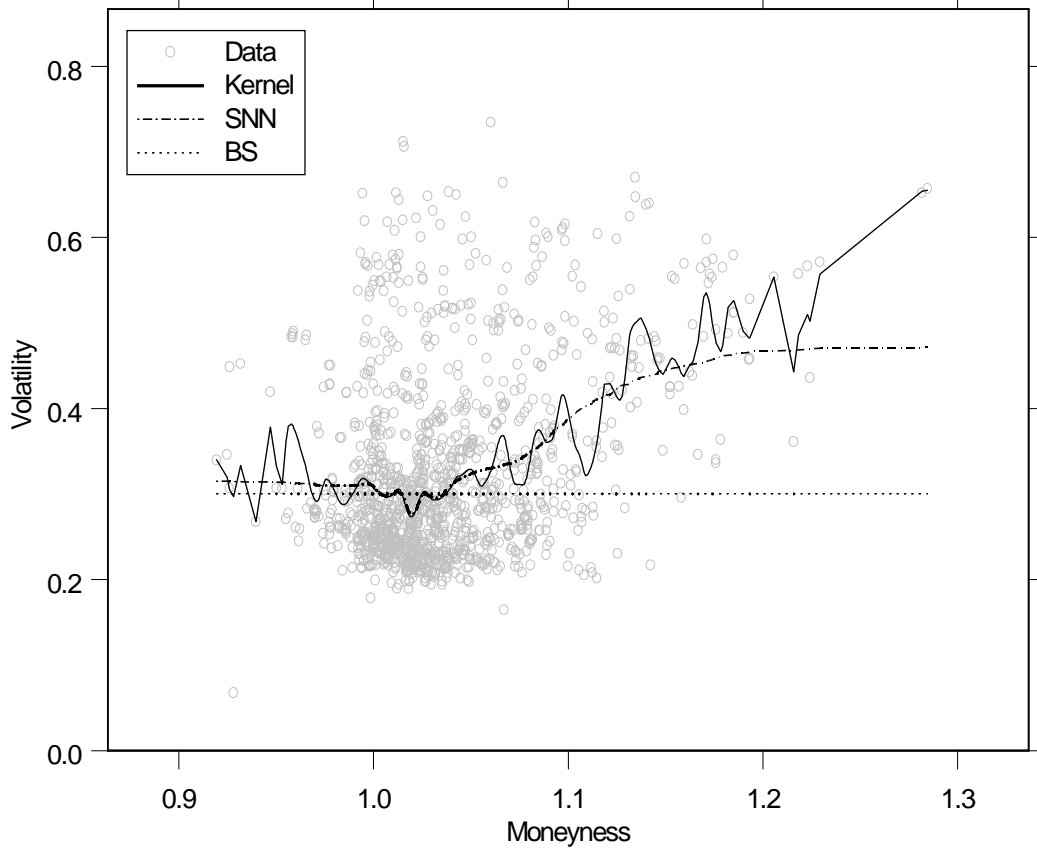


Figure 2: Nonparametric Volatility. Calls and Puts in 1998

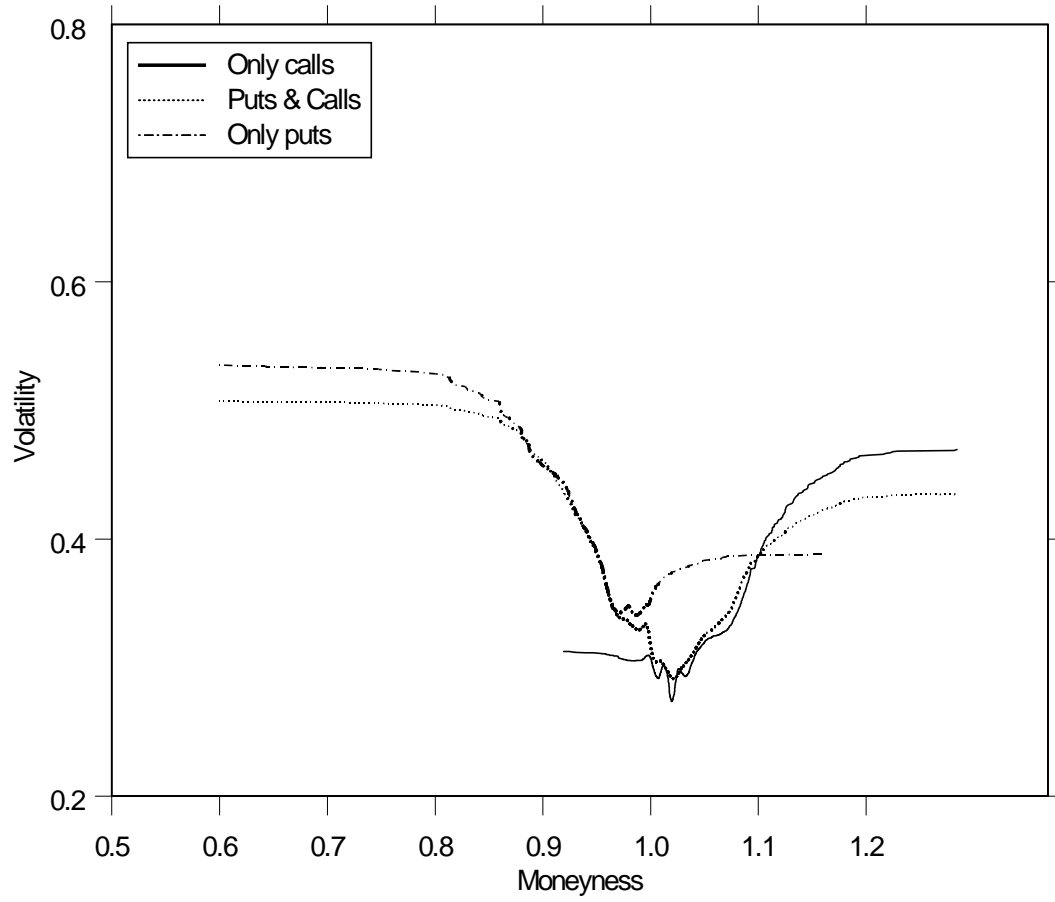


Figure 3: Risk-Neutral Densities for Calls and Puts in 1998: Liquidity Effects

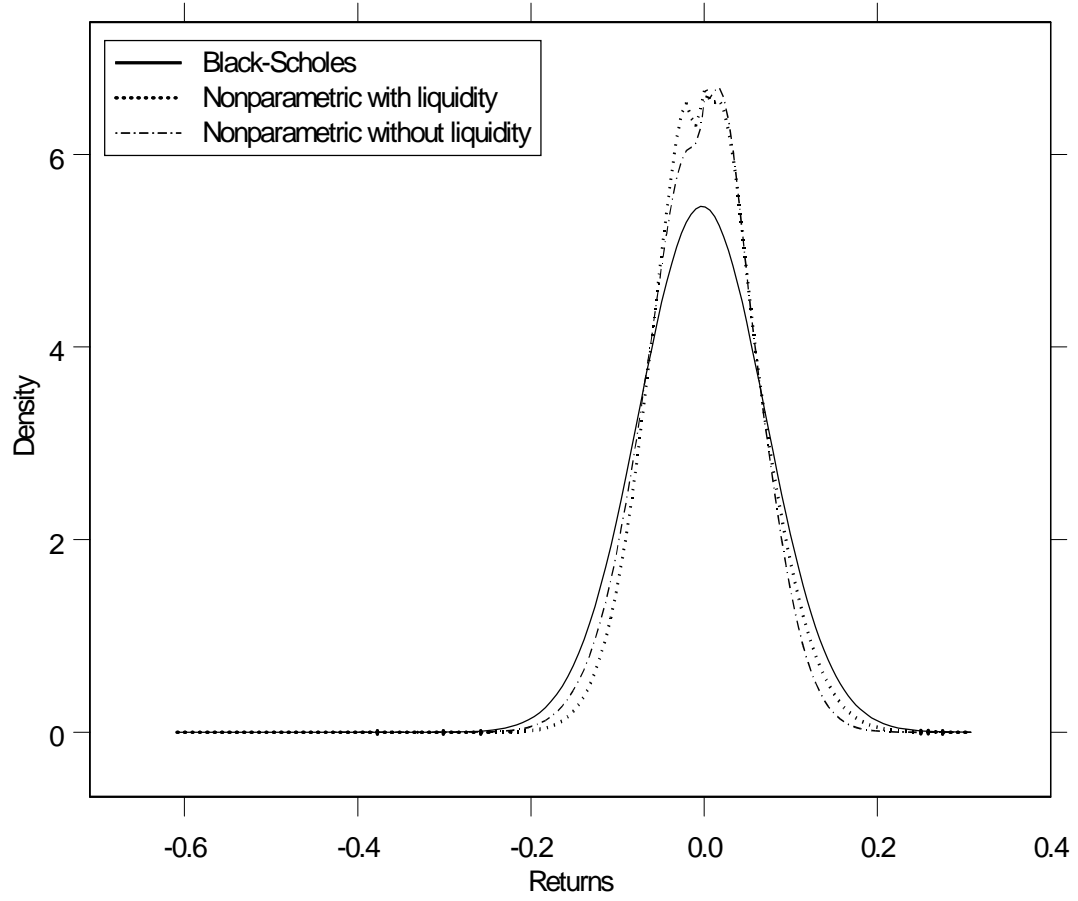


Figure 4: Risk-Neutral Densities for Calls in 1998

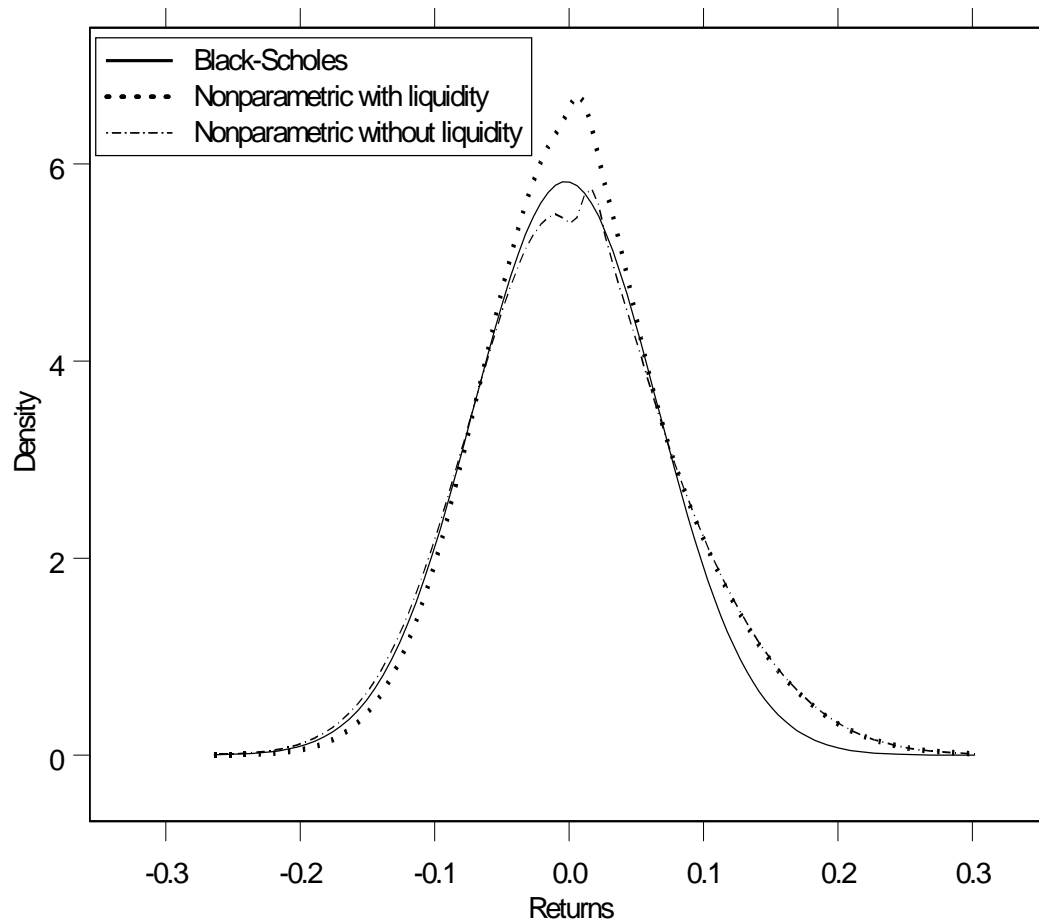


Figure 5: Risk-Neutral Densities for Puts in 1998

