LSF Research Working Paper Series

N°. 14-06

Date: March 2014

Title: Evaluating Option Pricing Model Performance Using Model Uncertainty

Author(s)*: Dennis Bams, Gildas Blanchard, Thorsten Lehnert

Abstract: The objective of this paper is to evaluate option pricing performance on the cross sectional level. For this purpose, we propose a statistical framework, in which we in particular account for the uncertainty associated with the reported pricing performance. Instead of a single figure, we determine an entire probability distribution function for the loss function that is used to measure option pricing performance. This methodology enables us to visualize the effect of parameter uncertainty on the reported pricing performance. Using a data driven approach, we confirm previous evidence that standard volatility models with clustering and leverage effects are sufficient for the option pricing purpose. In addition, we demonstrate that there is short-term persistence but long-term heterogeneity in cross-sectional option pricing information. This finding has two important implications. First, it justifies the practitioner's routine to refrain from time series approaches, and instead estimate option pricing models on a cross-section by cross-section basis. Second, the long term heterogeneity in option prices pinpoints the importance of measuring, comparing and testing option pricing model for each cross-section separately. To our knowledge no statistical testing framework has been applied to a single cross-section of option prices before. We propose a methodology that addresses this need. The proposed framework can be applied to a broad set of models and data. In the empirical part of the paper, we show by means of example, an application that uses a discrete time volatility model on S&P 500 European options.

Keywords: option pricing, cross-section, estimation risk, parameter uncertainty, specification test, bootstrapping

JEL Classification: G12, C15

***Corresponding Author's Address:** Tel: +352 46 66 44 6941; Fax : +352 46 66 44 6835
E-mail address: Thorsten.lehnert@uni.lu

Evaluating Option Pricing Model Performance Using Model Uncertainty

Dennis Bams
Limburg Institute of Financial Economics (LIFE), Maastricht University,
P.O. Box 616, 6200 MD Maastricht, The Netherlands

Gildas Blanchard
Limburg Institute of Financial Economics (LIFE), Maastricht University,
P.O. Box 616, 6200 MD Maastricht, The Netherlands

Thorsten Lehnert*
Luxembourg School of Finance, University of Luxembourg,
4, rue Albert Borschette, L-1246 Luxembourg, Luxembourg

March 2014

Abstract

The objective of this paper is to evaluate option pricing performance on the cross sectional level. For this purpose, we propose a statistical framework, in which we in particular account for the uncertainty associated with the reported pricing performance. Instead of a single figure, we determine an entire probability distribution function for the loss function that is used to measure option pricing performance. This methodology enables us to visualize the effect of parameter uncertainty on the reported pricing performance. Using a data driven approach, we confirm previous evidence that standard volatility models with clustering and leverage effects are sufficient for the option pricing purpose. In addition, we demonstrate that there is short-term persistence but long-term heterogeneity in cross-sectional option pricing information. This finding has two important implications. First, it justifies the practitioner's routine to refrain from time series approaches, and instead estimate option pricing models on a cross-section by cross-section basis. Second, the long term heterogeneity in option prices pinpoints the importance of measuring, comparing and testing option pricing model for each cross-section separately. To our knowledge no statistical testing framework has been applied to a single cross-section of option prices before. We propose a methodology that addresses this need. The proposed framework can be applied to a broad set of models and data. In the empirical part of the paper, we show by means of example, an application that uses a discrete time volatility model on S&P 500 European options.

Keywords: option pricing, cross-section, estimation risk, parameter uncertainty, specification test, bootstrapping.
JEL-Classification: G12, C15

Corresponding Author. Email: thorsten.lehnert@uni.lu

Evaluating Option Pricing Model Performance Using Model Uncertainty

Abstract

The objective of this paper is to evaluate option pricing performance on the cross sectional level. For this purpose, we propose a statistical framework, in which we in particular account for the uncertainty associated with the reported pricing performance. Instead of a single figure, we determine an entire probability distribution function for the loss function that is used to measure option pricing performance. This methodology enables us to visualize the effect of parameter uncertainty on the reported pricing performance. Using a data driven approach, we confirm previous evidence that standard volatility models with clustering and leverage effects are sufficient for the option pricing purpose. In addition, we demonstrate that there is short-term persistence but long-term heterogeneity in cross-sectional option pricing information. This finding has two important implications. First, it justifies the practitioner's routine to refrain from time series approaches, and instead estimate option pricing models on a cross-section by cross-section basis. Second, the long term heterogeneity in option prices pinpoints the importance of measuring, comparing and testing option pricing model for each cross-section separately. To our knowledge no statistical testing framework has been applied to a single cross-section of option prices before. We propose a methodology that addresses this need. The proposed framework can be applied to a broad set of models and data. In the empirical part of the paper, we show by means of example, an application that uses a discrete time volatility model on S&P 500 European options.

1. INTRODUCTION

Explaining option prices has been at the center of financial research interest since Black and Scholes (1973). Two complementary streams of research have emerged. The first one is dedicated to theoretically developed models that reflect complex data generating processes of the volatility dynamics. The second one, more pragmatic, is devoted to assessing, evaluating and comparing the performance of these alternative models empirically.

Our focus is on the latter. In the following we suggest a method to derive the probability distribution of a loss function. This allows for the statistical comparison of alternative models' ability to price a single cross-section of option prices, either in-sample or out-of-sample. Different from existing methods the proposed test does not rely upon a time series of cross-sections, nor is it limited to the comparison of absolute pricing errors alone.

How to measure option pricing model performance depends on the user's perspective. For instance, Bakshi, Cao and Chen (1997) provide three performance criteria, using alternatively internal consistency, pricing error and hedging error as objective evaluation criteria. Pricing error has become the leading criterion across a wide range of studies. The distance between observed and predicted options prices forms the basis of loss function based measures. This has the merit of an intuitive economic interpretation of a model's ability to match observed data. Moreover, this criterion allows for an out-of-sample comparison of models that are very different in nature.

The existing option pricing literature provides some critical insights on what model specifications and what estimation and evaluation criteria to use. Hardle and Hafner (2000) and Heston and Nandi (2000) show that adding a leverage effect to the standard autoregressive volatility model leads to a significant reduction of the option pricing error. Furthermore, Christoffersen and Jacobs (2004a) provide empirical evidence that the use of a specification which is richer than a standard asymmetric GARCH model does not provide any further pricing performance improvement. Nonetheless, Barone-Adesi, Engle and Mancini (2008) establish that additional pricing error reduction can be achieved by allowing for non-normality specifications of the error term.

Aside from model selection, major developments have been made on the methodological aspects of parameter estimation of option pricing models. Volatility dynamics, which are inferred from information in the underlying time series returns, differ from the volatility dynamics that result from a calibration on options prices directly (Engle and Mustafa, 1992). While initially studies have relied on parameter estimates under the physical measure (Bollerslev and Mikkelsen, 1999), it subsequently turned out that parameters estimated under the risk neutral measure better explain empirical features of option prices (Duan, 1996) and significantly reduce the value of the loss function (Christoffersen and Jacobs, 2004a). Accordingly, estimating parameters on options prices directly has become the standard practice (Lehnert (2003), Barone-Adesi, Engle and Mancini (2008), and Frijns, Lehnert and Zwinkels (2010)).

While practitioners prefer to continuously recalibrate their model on daily data, academics have suggested calibrating the model only once (Hull and Suo, 2002). Although less theoretically founded, the practitioners approach has proven to deliver better pricing performance (Christoffersen and Jacobs, 2004a) leading to a general adoption of this technique in academia too.

Despite the important insights in the existing literature regarding option pricing performance, we feel that there is an opportunity for improvement in the performance evaluation criterion, which is usually limited to a comparison of point estimates of loss functions. Our proposed performance evaluation criterion explicitly incorporates the effect of measurement, model and parameter uncertainty. Following the bootstrapping approach in Bams, Lehnert and Wolff (2009), we adopt an entire probability distribution function for the loss function, which facilitates a formal specification test to compare alternative models and approaches.

Since the literature advocates for a continuous recalibration, this stretches the need of comparing model performance at a cross-section by cross-section level. Our empirical results confirm the fundamental difference in nature between alternative cross sections of options. This heterogeneity in cross-sections of option prices over time makes statistical tests as proposed by Diebold and Mariano (1995) and applied for option pricing purposes in Christoffersen and Jacobs (2004a) less useful. This is a further motivation why to employ statistical inference at a single cross-section as proposed in our testing framework.

In the following we introduce our statistical framework and provide an empirical application of the framework, with a discrete time asymmetric GARCH model on S&P 500 options.

2. THE ECONOMETRIC FRAMEWORK

The statistical testing framework to measure option pricing model performance is general in the sense that it can be applied to any class of models such as continuous time models, discrete time models or ad hoc models. We present the framework in the context of a specific discrete time asymmetric GARCH volatility specification, to allow for realistic empirical findings and implications.

The choice of the discrete time asymmetric GARCH volatility specification is motivated by the evidence in the literature that such a model describes option pricing data features relatively well. Within the discrete time volatility model class, Christoffersen and Jacobs (2004a) show that, in order to obtain good pricing performance, clustering and leverage are two important effects to account for. In addition to clustering and leverage, other features have been proposed in the literature, but none of them have turned out to be substantially effective for additional pricing performance. Hsieh and Ritchken (2005) show that GARCH models are able to explain a significant portion of the volatility smile. Lehar, Scheicher, and Schittenkoph (2002) demonstrate the relative outperformance of GARCH models compared to stochastic volatility models in term of out-of-sample options pricing performance. We use the volatility model of Frijns, Lehnert and Zwinkels (2010). This discrete-time specification is one of the numerous available asymmetric GARCH models, that features both volatility clustering and volatility leverage.

Discrete-time volatility models including the estimation of the long term unconditional volatility parameter, are known to be unstable. Therefore, we have chosen to approximate the long term unconditional volatility in the model with the realized long term volatility. This approach is similar to variance targeting, and has proven to stabilize the volatility process at no pricing performance costs (Bams, Lehnert, and Wolff (2009)).

The continuously compounded returns of the underlying asset follow the traditional process:

$$r_t = \mu + \sigma_t \varepsilon_t \tag{1}$$

$$\varepsilon_t \sim N(0,1) \tag{2}$$

The volatility dynamics are defined as follows:

$$\mathrm{Ln}(\sigma_{t+1}^2) = \mathrm{Ln}(\sigma_t^2) + \frac{1}{2}\alpha(\sigma_t^2 - \overline{\sigma_t^2}) + \frac{1}{2}\{\beta_0 \max\{\varepsilon_t, 0\} - \beta_1 \min\{\varepsilon_t, 0\}\} \tag{3}$$

Volatility is a function of two equally weighted components. The first component, $\{\alpha(\sigma_t^2 - \overline{\sigma_t^2})\}$, drives the mean reversion of volatility where α is the parameter that determines the speed of mean-reversion and $\overline{\sigma_t^2}$ is the long-term unconditional volatility. The second component, $\{\beta_0 \max\{\varepsilon_t, 0\} - \beta_1 \min\{\varepsilon_t, 0\}\}$, features the leverage effect where β_0 and β_1 allow for an asymmetric response, to positive and negative shocks, respectively.

Our pricing performance results are in line with Barone-Adesi, Engle and Mancini (2008) who use a similar sample and another asymmetric GARCH specification. This provides comfort that the model used is representative for a wider class of asymmetric GARCH-type model.

Following Duan (1995), we apply the Local Risk Neutral Valuation Relationship to arrive at the return dynamics under the risk adjusted probability measure. The Local Risk Neutral Valuation Relationship specifies that the one period forward conditional variance is the same under both the actual and risk adjusted dynamics. For the conditional expectation of the underlying under the risk neutral probability is holds that:

$$E^{Q}[\exp(r_t)|\Omega_{\mathrm{t-1}}] = \exp(r_t^f) \tag{4}$$

where r_t^f is the risk free rate at time t. This results in the following risk adjusted process for the return dynamics:

$$r_t = r_t^f - \frac{1}{2}\sigma_{\mathrm{t}}^2 + \sigma_{\mathrm{t}}\varepsilon_{\mathrm{t}} \tag{5}$$

$$\varepsilon_t|\Omega_{\mathrm{t-1}} \sim N(0,1) \tag{6}$$

The volatility dynamics, as given in equation (3), remain unchanged when transitioning from actual to risk-adjusted return dynamics.

From the Principle of Risk Neutral Valuation, it follows that option prices are determined as the expected option payoff function discounted with the risk free rate, where expectations are taken under the risk adjusted probability measure, Q:

$$C_{it} = \exp(-r^f\left(t, T_{it}^C\right) \times \left(T_{it}^C - t\right)) \times E^Q\{\max[S\left(T_{it}^C\right) - K_{it}^C, 0]|\Omega_{\mathrm{t-1}}\} \tag{7}$$

$$P_{it} = \exp(-r^f(t, T_{it}^P) \times (T_{it}^P - t)) \times E^Q\{\max[K_{it}^P - S(T_{it}^P), 0]|\Omega_{\mathrm{t-1}}\} \tag{8}$$

Where C_{it} and P_{it} are, respectively, the call i and put i prices at time t; T_{it}^C and T_{it}^P are the associated times-to-maturity of call i and put i at time t; K_{it}^C and K_{it}^P are the strike prices of the call and put options i at time t. With $r^f(t,T)$ we indicate the risk free rate at time t, appropriately reflecting the term structure of interest rates for the remaining time-to-maturity, $(T-t)$. Finally, $S(T)$ is the value of the underlying stock at time T.

In the absence of closed or semi-closed form solutions, the option payoff distribution in equations (7) and (8) is obtained through Monte Carlo Simulation. Following Duan and Simonato (1998), we use the Empirical Martingale Simulation (EMS) approach to reduce the required number of simulations for convergence.

Subsequently, cross sectional parameter estimation follows by the choice and subsequent minimization of a loss function that calibrates modeled option prices to observed option prices. The proposed statistical framework is general in the sense that it works for alternative loss functions. Following the recommendation of Bams, Lehnert, Wolff (2009), for the empirical application the root mean squared error (RMSE) of absolute pricing errors is chosen as loss function. Parameter estimation for cross-section *t*, follow from minimization of the following loss function:

$$RMSE_t = \sqrt{\frac{1}{N_t^C + N_t^P}\left(\sum_{i=1}^{N_t^P}(P_{it} - \hat{P}_{it})^2 + \sum_{i=1}^{N_t^C}(C_{it} - \hat{C}_{it})^2\right)} \tag{9}$$

where N_t^C and N_t^P are the respective number of call and put options in cross-section t; $\hat{C}_{it}$ and $\hat{P}_{it}$ are the model call i and put i prices at t, following from equations (5) to (8); C_{it} and P_{it} are the observed call i and put i options prices at time t. The loss function in equation (9) is minimized for each cross-section of option prices *t* separately using the Newton-Raphson algorithm, resulting into a separate set of parameters estimates $(\widehat{\alpha}, \widehat{\beta_0}, \widehat{\beta_1})_t$ as well as an accompanying value for the loss function, $\widehat{RMSE}_t$ for each cross section *t*.

The absence of an analytical or easily derivable distribution function for the loss function causes the lack of statistical inference in the majority of the option pricing performance evaluation literature. Comparison of absolute loss functions for model selection is characterizing the option pricing model literature, instead. Bakshi, Cao and Chen (1997) benchmark the alternative continuous time models based on the absolute RMSE; Heston and Nandi (2000) use the RMSE accompanied by other absolute loss functions to demonstrate the outperformance of their model over available alternative specifications; Barone-Adesi, Engle, and Mancini (2008) use a battery of absolute loss functions to depict the pricing improvement of a model that includes historical innovations compared to Gaussian or other parametric alternatives.

Christoffersen and Jacobs (2004a), provide more statistical validity to option pricing model comparison using the Diebold-Mariano (DM) test. This test is virtually a z-test on a time series differential between two forecasts losses corrected for serial correlation. In option pricing, the loss function differential time series is obtained by iterative cross section by cross section estimation and loss function evaluation pooled together over all cross sections. The DM test is designed for out-of-sample forecast comparison, which excludes in-sample testing. Moreover, the DM test is not intended for model selection (Diebold, 2012). In our opinion, the principal limitation of the DM test, resides in its inability to address the model performance for a single cross section.

The time series requirement, pooling the information in many cross-sections, would at best be an approximation for the model performance at a single cross section. This would require homogeneity of the information in the alternative cross-sections. In fact, there are two sources of heterogeneity, as we will also demonstrate in the empirical application. First, model performance is affected by the change in the economic environment. Diebold and Mariano (1995) discuss the effect of business cycles on relative predictability. Second, the iterative recalibration comes at a cost. The data contained in each cross section is evolving through time.

The total number of observations and the qualitative composition of cross sections vary significantly. This suggests that over time the informational content of alternative cross sections is changing and therefore different.

To answer the need for a cross sectional test, we propose a bootstrap based methodology to estimate the distribution of each cross-section RMSE separately. In the general finance literature, the use of bootstrap to overcome the lack of analytical solutions is a common practice (Bams, Lehnert and Wolff (2005); Ledoit and Wolf (2008); Hansen, Lunde and Nason (2011)). However, the use of bootstrapping in the specific field of option pricing has been more limited. Christoffersen and Jacobs (2004b) use a jackknife approach to study the pricing effect of alternative loss functions on a contemporaneous out-of-sample observation. Bams, Lehnert and Wolff (2009) bootstrap and summarize the loss function distribution into a statistic to assess loss function selection accounting for uncertainty. Finally, in a simulation and application study, Yatchew and Härdle (2006) show in the context of nonparametric state price density estimation that relying on a wild bootstrap to construct call function confidence intervals leads to reasonable results. We take these evidences as good indication that bootstrapping option pricing errors is an appropriate method to depict pricing uncertainty.

Bootstrapping is a re-sampling technique to obtain the distribution of a particular statistic. In the following, this method is applied to obtain the probability distribution of the loss function for a single cross-section. The initial estimation step produces parameter estimates, fitted option prices, residuals and accompanying value for the loss function. The difference between the observed market price and theoretical price is the residual. We introduce the following matrix notation:

$$C_t \equiv \left(C_{it}, \dots, C_{N_t^C,t}\right)' \qquad t = 1, \dots, T \tag{10}$$

$$P_t \equiv \left(P_{it}, \dots, P_{N_t^P,t}\right)' \qquad t = 1, \dots, T \tag{11}$$

where C_t and P_t are two vectors respectively (N_t^CX 1) and (N_t^P X 1) of the observed call and put market prices; $\hat{C}_t$and $\hat{P}_t$ are two vectors respectively (N_t^CX 1) and (N_t^P X 1) of the modeled call and put prices:

$$\hat{C}_t \equiv \left(\hat{C}_{it}, \dots, \hat{C}_{N_t^C,t}\right)' \qquad t = 1, \dots, T \tag{12}$$

$$\hat{P}_t \equiv \left(\hat{P}_{it}, \dots, \hat{P}_{N_t^P,t}\right)' \qquad t = 1, \dots, T \tag{13}$$

We introduce two vectors of residuals:

$$\eta_t^C \equiv \left(\eta_{it}^C, \dots, \eta_{N_t^C,t}^C\right)' \qquad t = 1, \dots, T \tag{14}$$

$$\eta_t^P \equiv \left(\eta_{it}^P, \dots, \eta_{N_t^P,t}^P\right)' \qquad t = 1, \dots, T \tag{15}$$

where it holds that:

$$\eta_{it}^C = C_{it} - \hat{C}_{it} \qquad i = 1, \dots, N_t^C \qquad t = 1, \dots, T \tag{16}$$

$$\eta_{it}^P = P_{it} - \hat{P}_{it} \qquad i = 1, \dots, N_t^P \qquad t = 1, \dots, T \tag{17}$$

It is possible to create a bootstrapped sample by constructing “bootstrapped market prices” for each observations drawing residual with replacement. $C_t^*$and P_t^* are two vectors respectively (N_t^CX 1) and (N_t^P X 1) of the bootstrapped call and put prices, defined as:

$$C_t^* \equiv \left(C_{it}^*, \dots, C_{N_t^C,t}^*\right)' \qquad t = 1, \dots, T \tag{18}$$

$$P_t^* \equiv \left(P_{it}^*, \dots, P_{N_t^P,t}^*\right)' \qquad t = 1, \dots, T \tag{19}$$

where B_t^C and B_t^P are two vectors respectively (N_t^C X 1) and (N_t^P X 1) obtained from drawing with replacement N_t^C and N_t^P observations from the η_{it}^C and η_{it}^P vectors, defined as:

$$B_t^C = (B_{it}^C, \dots, B_{N_t^C,t}^C)' \qquad t = 1, \dots, T \tag{18}$$

$$B_t^P = (B_{it}^P, \dots, B_{N_t^P,t}^P)' \qquad t = 1, \dots, T \tag{19}$$

It holds that:

$$C_t^* = \hat{C}_t + B_t^C \qquad t = 1, \dots, T \tag{20}$$

$$P_t^* = \hat{P}_t + B_t^P \qquad t = 1, \dots, T \tag{21}$$

The drawing procedure to obtain equation (20) and (21) can be replicated independently S times to obtain S bootstrapped samples. The optimization of equation (9) is performed on each of the S bootstrapped samples for cross section t. This mechanism procures the desired distribution reflecting estimation uncertainty of the estimated parameters $(\hat{\alpha}, \hat{\beta}_0, \hat{\beta}_1)_t^{(1)}, \dots, (\hat{\alpha}, \hat{\beta}_0, \hat{\beta}_1)_t^{(S)}$ and loss functions $\widehat{RMSE}_t^{(1)}, \dots, \widehat{RMSE}_t^{(S)}$.

The distribution of the residuals in different maturity and moneyness categories is strongly diverging. The deep in-the-money long maturity residuals are in absolute (relative) term larger (smaller) and more (less) volatile than deep out-of-the-money short maturity residuals. To account for this cross-sectional heterogeneity across moneyness and maturity, a block bootstrapping technique is pursued as applied in Bams, Lehnert and Wolff (2009). Bootstrapped prices are formed by drawing residuals from the block matching the moneyness/maturity block of the particular option. For that purpose, similar to Barone-Adesi, Engle, and Mancini (2008), 12 blocks are formed with respect to maturity ($60 > T, 60 \leq T \leq 160, T > 160$) and moneyness ($0.85 > M, 0.85 \leq M \leq 1, 1 < M < 1.15, M > 1.15$). To have a sufficient large historical sample, we bootstrap residuals from the contemporaneous and the 3 previous cross sections. This procedure avoids a downward bias in the bootstrapped confidence interval due to relying too often on the same bootstrapped error terms, while short term homogeneity in option prices cross sections safeguards the appropriateness of drawing from this larger sample.

3. DATA

We use European S&P 500 index options (SPX). The SPX option market is the most active in the world, making it popular in the option pricing literature (Barone-Adesi, Engle, and Mancini (2008)) and allows for a comparison with findings in existing literature. The closing prices of each Wednesday are used. These data are collected from OptionMetrics and cover 3 years of data from January 2002 to December 2004, including 155 Wednesdays.

Filtering criteria are comparable to Barone-Adesi, Engle, and Mancini (2008). Only out-of-the-money options are selected. Options with a maturity lower than or equal to 10 days and higher than or equal to 360 days are filtered-out. Similarly options with an implied volatility higher than 70% and options with a price lower than or equal to $0.05 are excluded. In addition to these conventional rules, if in the same cross-section two options have the same maturity and strike price only the most traded option remains in the sample. The S&P 500 dividend yields and zero–coupon default free interest rates are also collected from OptionMetrics.

**** PLEASE INSERT TABLE 1 HERE ****

Table 1 presents the average price, average implied volatility and number of contracts per category, where a category is defined by moneyness and maturity. We observe the well-known characteristics of the volatility smile, that deep out-of-the-money puts and calls exhibit a higher implied volatility than close to the money put and call options. This difference in implied volatility decreases with maturity. The division of options within a moneyness/maturity category is balanced with respect to proportion of puts and calls. Put (call) options represent 51% (49%) of the sample; 48% of the options are deep out-of-the-money with a strike/underlying ratio under 0.85 for puts and over 1.15 for calls; 52% of the options are less out-of-the-money with a strike/underlying ratio over or equal to 0.85 for puts and under or equal to 1.15 for calls. Long, medium and short maturity options represent respectively 36%, 31% and 33% of the sample.

To assess the sources and effects of cross-sectional heterogeneity, we collect descriptive statistics of market conditions and sample conditions at the individual cross-sectional level. Table 2 Panel A reports the average, standard deviation, minimum and maximum values for market conditions and cross-sectional composition variables. Cross-sections differ strongly with respect to both types of variables.

**** PLEASE INSERT TABLE 2 HERE ****

The selected sample period displays changing market conditions. From 2002 to 2004, the average yearly volatility is 16.6%. Some extremely high volatility regimes are observed at the beginning of the sample, reaching 54.1%. In both 2003 and 2004, the volatility is much lower; the minimum observed yearly volatility is 7.1%. The high volatility year 2002 coincides with a bearish market. The worst 3 months loss equals -23%, while in the subsequent year, the market increased by 25% within 3 months. These statistics highlight the significant change in the underlying physical and risk neutral probability measure across time.

Table 2 Panel B reports high correlations between market and sample composition variables. These correlations are mechanically magnified by the exclusion filtering rule used to select the included options in our sample. The sampling condition causing this cross-section composition market conditions dependence is the restriction to only include out-of-the money options. Under this condition, a turbulent market that is characterized as a period preceded by low return, high realized volatility and high VIX results in subsequent cross-sections that are over-dominated by slightly out-of-the-money call options. When a market crashes, the majority of the put options become in-the-money and are automatically excluded from the sample. The reverse occurs for previously in-the-money call options. These freshly included call options are concentrated in a close-to-money category changing the distribution of option moneyness in the cross-section.

The resulting composition shifts are potentially substantial. Call options representation in a cross section ranges between 24% and 81%. Hence, the balanced distribution of puts and calls found for the total sample often does not hold at the single cross-sectional level. The accompanying variation in average moneyness is also meaningful, which is important because options with different levels of moneyness carry different information regarding the risk neutral distribution. Average moneyness per cross section ranges between 13% and 35%. Average option prices per cross section take a minimum value of $6.80 and a maximum value of $16.78. The low average options prices are driven by an important concentration of "very cheap option". For some cross sections, we observe up to almost 40% of options with prices below $1. Absolute prices are relevant since we use the RMSE as loss function. The composition and distribution of option prices in a particular cross section have an implicit effect on the weight allocated to different observations and hence affect the estimation results.

4. Empirical Analysis

Parameter estimates - cross section by cross section

Table 3 presents summary statistics for the cross section by cross section estimation results of the volatility specification in equation (3), by minimizing the objective function in equation (9). The reported numbers are summarizing the individual results from 155 cross sectional parameter estimates.

**** PLEASE INSERT TABLE 3 HERE ****

The average in-sample RMSE ($1.12) is comparable with the Barone-Adesi, Engle, and Mancini (2008) in-sample statistic for the Heston Nandi model. This statistic confirms that our volatility specification has pricing performance similar to other asymmetric GARCH type models and fits the data sufficiently well. All coefficients display expected values accounting for the clustering and asymmetric dynamics as reflected by the positive β_1 coefficient.

These results pinpoint the heterogeneity over the alternative cross sections. Loss functions are highly fluctuating over time ranging between a minimum of $0.61 to a maximum of $2.38. There exist different regimes of pricing errors. Cross-sections of option relatively mispriced (well-priced) tend to be followed by other mispriced (well-priced) cross-sections. The observed heterogeneity over time, limits the possibility of pooling many cross sections to arrive at reliable test statistics, and supports instead our proposed block bootstrapping approach, where only a limited number of cross sections are pooled for inference purposes.

Appropriateness bootstrapping design

To get a sense of the appropriateness of our bootstrapping design, Figure 1 Panel A plots observed prices against the average of bootstrapped prices. Even with a limited number of repetitions, the mean bootstrapped prices match the observed data well, as illustrated by the plots being concentrated within a tight range along the diagonal line.

**** PLEASE INSERT FIGURE 1 HERE ****

Figure 1 Panel B compares the time series of cross-sectional RMSE resulting from the non-linear least squares optimization in (9) and the average of the accompanying bootstrapped RMSEs for each point in time. The average bootstrapped RMSE is tightly tracking the estimated RMSE, capturing the changes in economic regime. We interpret these findings as good indication of the reliability of our bootstrapping procedure.

We also investigate the properties of the bootstrap by a comparison of average bootstrapped prices and average observed prices at option category level (related to moneyness and maturity). Table 4 presents the results of a mean tests per category. For all categories, equality of means cannot be rejected.

**** PLEASE INSERT TABLE 4 HERE ****

Parameter inference – standard deviations

As a result of the cross sectional estimation and bootstrapping procedure, the distribution and variance-covariance of the parameters estimates are also naturally available at the single cross sectional level. Dumas, Fleming and Whaley (1998) used as an alternative the time series variations in continuously re-estimated parameters to investigate their dynamics and provide inference. We compare standard deviations and correlations for parameter estimates from both a time series based approach as in Dumas, Fleming and from a cross-sectional approach as resulting from the bootstrap approach. It turns out that bootstrapped parameters statistics are not only available on an individual cross sectional frequency, but they also contain different information.

**** PLEASE INSERT TABLE 5 HERE ****

Table 5 presents the time series based standard deviations as well as the average of cross-sectional standard deviations. The time series based standard deviations are remarkably larger than their cross-sectional counterparts. Results suggest a ratio of 3 to 5 times higher times series based standard deviations. The time series of standard deviations reflects two sources of variations. The first is the variation due to uncertain parameter estimates and over-fitting within each cross-section. The second is the variation across cross-sections driven by a change in economic conditions over time. The cross-sectional approach isolates the first source of variation from the second, which makes the standard deviations more reliable as a reflection of parameter uncertainty at the single cross sectional level.

The local volatility parameter, which is the parameter that is used as starting value for the volatility process in equation (3), shows a difference that is even more pronounced, with a ratio of 32. Local volatility tracks extremely tightly the economic conditions since it represents an instantaneous measure of market turbulences. The time period covered exhibits extremely quiet and turbulent market at time. The change in economic conditions explains the gap between the local volatility time series and cross-sectional statistic.

This evidence suggests that cross-sections strongly diverge in nature and require individual consideration. This finding is consistent with the argument that continuous recalibration is desirable.

Parameter inference – correlations

We next turn to the comparison of time series and cross-sectional based parameters variance-covariance matrix. The left panel of table 6 exhibits the time series correlations while the top right panel presents the average cross-sectional correlations. The minimum and maximum correlations are also available for the cross-sectional case, in the lower right panel. The time series and cross-sectional correlations have the same signs, yet the magnitudes are significantly different. The cross-sectional correlations are extremely high, while the time series based correlations are driven down by the noise created by the second source of variation discussed in the previous paragraph.

**** PLEASE INSERT TABLE 6 HERE ****

At the cross-sectional level, the parameters β_0 and β_1 are almost perfectly negatively correlated, with a correlation coefficient of minus one. This suggests that the relative difference between β_0 and β_1 is more relevant to capture asymmetry rather than the values taken by these two parameters. Hentschel (1995) shows that volatility asymmetry can be modeled either as a shift or a rotation parameter. On the one hand, a shift implies that the conditional volatility response to negative shocks is higher than to positive shocks by a constant factor. On the other hand, rotation suggests that the conditional volatility response is more complex and requires two different slopes for responses to either positive or negative shocks. Our results shows that, for option pricing purposes, the shift parameter is the main driver of volatility asymmetry.

The cross-sectional correlation results suggest that the mean reversion parameter α and parameters β_0 and β_1 are also remarkably highly correlated. The dependence on past conditional volatility and past shocks are competing to generate the clustering effect needed to match the data. In a traditional GARCH (1,1) sense, this implies that the pricing performance would not be sensibly altered for alterative parameter values as long as the sum of these two parameters are close to one.

We interpret these substantially high cross-sectional correlations as evidence that the loss function surface is flat, with many alternative local optima that are close to the global optimum. This suggests that a simple volatility specification is already over-fitting the data. Dumas et al. (1998) used the difference between in- and out-of-sample pricing errors to prove that simple ad hoc Black and Scholes models are over-fitting too. Our results show that this problem also exists for more theoretically founded discrete time volatility models. These findings therefore favor the use of a parsimonious volatility specification rather than richer specification prone to over-fitting. This recommendation is aligned with the relative good performances obtained by an asymmetric GARCH compared to richer specifications in Christoffersen and Jacobs (2004a).

Inference - loss distribution

The loss function distribution function is relevant to assess option pricing model performance. The outcome of the bootstrapping approach is an entire probability distribution of the loss function, i.e. RMSE, for each cross-section. The variation in RMSE is unaffected by the heterogeneity between cross-sections and is specific to the individual nature of the particular cross-section.

**** PLEASE INSERT FIGURE 2 HERE ****

Figure 2 Panel A illustrates graphically the RMSE distributions obtained for three particular chosen cross-sections. The distribution around the RMSE point estimate can be truly wide as it is suggested by the cross-section of 2002/5/8. This illustrates that the mean of the distribution is not sufficient information for a pricing performance evaluation, and instead a statistical test based on the entire distribution is warranted.

The mean of the RMSE distribution is not the only statistic that changes between cross-sections. The shape of the distribution functions also diverges strongly. Panel C shows, by means of example, the discrepancy of the RMSE distribution shapes for two cross-sections with an equivalent RMSE level. Table 7 present the results of Kolomogrov-Smirnov tests confirming that our visual inspection is correct.

**** PLEASE INSERT TABLE 7 HERE ****

We conclude that cross-sections display significant different levels of RMSE and diverging RMSE distribution shapes, providing additional evidence that a statistical test based on the entire distribution is preferred over a comparison of averages alone.

Sources of time variation in RMSE

The previous results highlight the time varying nature of the RMSE distribution. Next, we investigate the time dynamics of the RMSE distribution in a simple time series regression framework. For this analysis a single measure to describe each RMSE distribution is needed. For this purpose we use three different representations, being the mean of the RSME distribution, the coefficient of variation and the Asymmetric Selection Criterion (ASC), as defined in Bams, Lehnert, Wolff (2009):

$$ASC = -\frac{1}{\overline{RMSE}} \times \frac{\overline{RMSE} - F_{2.5\%}^{-1}(RMSE)}{\overline{RMSE} - F_{97.5\%}^{-1}(RMSE)} \tag{22}$$

where $\overline{RMSE}$ is the mean, $F_{2.5\%}^{-1}(RMSE)$ and $F_{97.5\%}^{-1}(RMSE)$ are the respective percentiles of the bootstrapped vector $RMSE$. The ASC statistic reflects a preference of negatively skewed loss distributions, implying below-average mispricing, to positively skewed counterparts. Moreover, the ASC penalizes high average and above-average RMSE while rewards below-average RMSE.

Table 8 reports the results of time series regressions investigating the factors influencing RMSE's distribution. The three proposed measures for the RMSE distribution are regressed on three types of explanatory variables. Lagged values are used to investigate the persistence of the loss function distribution. The local volatility parameter and the leverage parameter capture the effect of market conditions. The local volatility is equivalent to the VIX. We use the parameter estimates for β_1 in equation (3) as a proxy for the risk neutral distribution skewness. Both variables are known to be measures of market turbulence and fear. The average price, the proportion of very cheap options and the average maturity capture the effect of the sample composition.

**** PLEASE INSERT TABLE 8 HERE ****

The lags are the most powerful explanatory variables. Lagged ASC and lagged coefficient of variation standalone explain respectively about 60% and 70% of the variation, as indicated by the reported R-squares for the autoregressive regressions. Moreover, the one period lagged coefficient for ASC is equal to 0.84, which suggest persistence in the short term, suggesting that the RMSE distribution of a particular cross-section is a reliable source of information for the distribution of next week's RMSE.

Market conditions affect the model pricing performance to some extend as well. The local volatility and the leverage have a significant positive effect on the change in RMSE and coefficient of variation. The leverage variable has a significant negative effect on the ASC. The results suggest that turbulent markets characterized by high volatility and negatively skewed risk neutral distributions are accompanied by important mispricing as well as magnified uncertainty around the RMSE. This is in line with the well-known fact that GARCH models cannot account fully for the risk neutral skewness (Barone-Adesi, Engle, and Mancini (2008)) and the CBOE VIX (Hao and Zhang, 2013). Therefore, high risk neutral skewness and VIX are coupled with a high unexplained portion of the distribution leading to mispricing.

The effect of including sample composition variables is limited because of the important multi-collinearity with the market condition variables. The regression still provides interesting insights regarding the role of sample composition on the loss function distribution. After controlling for the local volatility, the average price of a cross-section seems to slightly reduce the mispricing and the coefficient of variation. This would mean that higher prices are easier to match and to predict more precisely. However, an unconditionally higher average price implies higher volatility and worst pricing performances. A high average price is associated with high mispricing because RMSE is an absolute measure. Extremely low priced options are associated with higher RMSE distribution dispersion.

Sample composition effects

Cheap options are often the very short maturity deep out-of-the money options. In order to change the distribution of the payoff to match the price of these options, volatility dynamics should be forced to have unrealistic, unstable parameters that would not match long term maturity options. The choice of a loss function (RMSE) targeting the absolute and not the relative pricing error will mechanically leads to disregard cheap options. Even if these options are completely 100% mispriced, this will in absolute terms still appear as a good performance driving the RMSE down. More weight is given to match longer term and more expensive options.

**** PLEASE INSERT FIGURE 3 HERE ****

As a result, in case of a loss function that uses RMSE as evaluation criterion, the informational value of cheap options is almost non-existent, while simultaneously leads to higher pricing uncertainty. Figure 3 explicitly pictures this effect. A high dispersion is displayed for average option prices. Certain cross-sections' average option prices are lower than $7. These low values can be the result of a high loading in inexpensive options.

The left side of the graph displays a clear negative relationship between uncertainty and average option prices. However, after a threshold around $10 this relationship disappears. This is evidence that the lowest average price cross-sections were loaded with these inexpensive options. After the threshold, we can assume that the relationship does not hold because the average price variation is not driven by the quantity of extremely inexpensive options. These results motivate the exclusion of very cheap options and the adoption of an alternative filtering rule that allows for more options to be selected in order to have enough options per cross-section.

Test example – local volatility parameter

To illustrate this model pricing performance comparison procedure, we introduce a second model, which is the model with the same volatility dynamics as in (3), while the local volatility parameter is not treated as an unknown parameter, but is fixed to the 14 days realized historical variance. Evaluating the pricing benefit of local volatility estimation is an interesting application of our framework for two reasons. First, this has not been widely studied in the previous literature (Bams, Lehnert, Wolff (2009)) and most empirical applications are silent about how they treat local volatility. Second, as suggested by previously reported parameter inference, local volatility is highly dependent on market conditions. Hence, the benefit of estimating local volatility is also likely to be time-varying.

The bootstrapped confidence intervals of both models' RMSE are used to test whether the unconstrained model has a statistically significant better pricing performance than the constrained model. Our approach is slightly time dependent since the residuals are drawn from the contemporaneous and the three previous cross-sections. To acknowledge for this limitation we, not only estimate the simpler specification, but also bootstrap it and use the mean bootstrapped RMSE as a reference for testing. This approach takes into account that, for cross-sections preceded by important pricing performance changes, the estimated RMSE and the bootstrap mean RMSE can diverge.

**** PLEASE INSERT FIGURE 4 HERE ****

Figure 4 draws the two models' RMSE confidence intervals. The lighter grey interval represents the distribution of the constrained model's RMSE. The darker grey interval represents the distribution of the unconstrained model's RMSE. The darkest areas indicate when the two distributions overlap. Our statistical test shows that estimating local volatility significantly reduces pricing error. The underlying time-series backward looking information is not a good substitute to the forward looking risk-neutral information relevant for option pricing. In 78% of the cross-section the two RMSE's confidence intervals are distinct enough to conclude that the unconstrained model outperforms significantly at the 5% level the constrained model. However, this leaves 12% of cross-sections where the two models RMSE point estimates are within the other model´s confidence intervals. Estimating the local volatility does not generate a statistical significant pricing improvement for these cross-sections. Solely comparing the RMSE point estimates for these cross-sections would naturally lead to favor the unconstrained model. Inspections of the whole RMSE distribution offers a diverging picture stretching the importance of accounting for uncertainty.

The number of failure to reject is substantial. Additionally these failures are not randomly distributed across time but they appear clustered. These clusters of overlapping RMSE distributions are noticeable in the middle of 2002 and at the end of 2004. This indicates that there are times were the two models pricing performance cannot be distinguished. Interestingly, the failures to distinguish the two models' pricing performances are concentrated during market stressed period with high volatility. The global context appears to influence the significance of the results. These times can be interpreted as either periods, where the cross-sectional data is not sufficient to precisely estimate the local volatility parameter or periods, where the economic conditions suggest that the local volatility converges to the realized volatility.

Both explanations diverge but these results are additional evidence that cross-sections are heterogeneous and require individual treatment. For that reason, option pricing models need to be evaluated, tested and benchmarked with respect to a certain informational and economic context. Our bootstrapping approach to construct confidence intervals allows for such a comparison in a rigorous statistical manner. For instance, the conclusions of the cross-sectional tests are opposites for end of 2002 and for end of 2003 cross-sections. A test using multiple cross-sections from 2002 and 2003 would have yielded mixed and inconclusive results.

5. CONCLUSION

Altogether, the empirical evidences presented in this paper suggest that cross-sections of option need to be regarded as independent entities. The long term cross-sectional heterogeneity affects both the pricing performance and the model selection risk. Option pricing models need to be benchmarked at the cross-sectional level to reflect the changing nature of the cross-section. Since cross-sections are similar on a short term horizon, this information is relevant for subsequent cross-sections. Nevertheless, one cannot pool too many cross-sections cannot without running the risk of being affected by the change in cross-sectional nature. Traditional option pricing statistical tests are inherently affected by this issue because of the important number of required cross-sections. Naturally, an absolute loss function comparison ignoring model selection risk and estimation uncertainty is not a better alternative.

The main novelty and innovation of our methodology is to provide a statistical framework to benchmark models at the cross-sectional level. The bootstraps result in a loss function distribution and confidence interval around the RMSE of one specific model. The statistical significance of alternative model pricing performance differences, accounting for model selection risk, is available for each cross-section. The empirical application of our framework yield interesting results.

Firstly, we confirm the result of Christoffersen and Jacobs (2004a), that discrete time volatility model including clustering and simple asymmetry effect fit option data well. Richer specifications are not recommended. We use in-sample bootstrapped variance-covariance matrices to show that a simple model is already over-fitting significantly.

Secondly, we quantitatively demonstrate the important heterogeneity in cross sections of options. Different cross sections strongly diverge in nature and cannot be considered identical. This time distinctness is highlighted by the discrepancy between time series information and cross-sectional bootstrapped information. Economic conditions, sample composition and information content are causing this diversity. These differences are aggravated with time since on a short term horizon cross sections can considered sufficiently homogeneous, which does not hold for longer horizons. These results encourage the practitioners' custom of continuous recalibration, profiting from the short term persistence, but allowing for longer term variations.

Thirdly, we show that the conclusion of a specification test is cross sectional dependent and cannot be generalized for all cross sections. This implies that the real question to ask is not what model is best, but what model is the best under what conditions. Our framework allows for an answer to this last question. The empirical application of this paper is limited to a specific type of discrete time volatility model, but the methodology is implementable for any class of option pricing model.

REFERENCES

Bakshi, G., Cao, C., & Chen, Z. (1997). Empirical performance of alternative option pricing models. *The Journal of Finance*, 52(5), 2003-2049.

Bams, D., Lehnert, T., & Wolff, C. C. (2005). An evaluation framework for alternative VaR-models. *Journal of International Money and Finance*, 24(6), 944-958.

Bams, D., Lehnert, T., & Wolff, C. C. (2009). Loss functions in option valuation: a framework for selection. *Management Science*, 55(5), 853-862.

Barone-Adesi, G., Engle, R. F., & Mancini, L. (2008). A GARCH option pricing model with filtered historical simulation. *Review of Financial Studies*, 21(3), 1223-1258.

Bollerslev, T., & Ole Mikkelsen, H. (1999). Long-term equity anticipation securities and stock market volatility dynamics. *Journal of Econometrics*,92(1), 75-99.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *The journal of political economy*, 637-654.

Christoffersen, P., & Jacobs, K. (2004). Which GARCH model for option valuation?. *Management Science*, 50(9), 1204-1221.

Christoffersen, P., & Jacobs, K. (2004). The importance of the loss function in option valuation. *Journal of Financial Economics*, 72(2), 291-318.

Diebold, F. X. (2012). *Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold-mariano tests* (No. w18391). National Bureau of Economic Research.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & economic statistics*, 13, 253-263

Duan, J. C. (1995). The GARCH option pricing model. *Mathematical finance*,5(1), 13-32.

Duan, J. C. (1996). Cracking the smile. *Risk*, 9(12), 55-59.

Duan, J. C., & Simonato, J. G. (1998). Empirical martingale simulation for asset prices. *Management Science*, 44(9), 1218-1233.

Dumas, B., Fleming, J., & Whaley, R. E. (1998). Implied volatility functions: Empirical tests. *The Journal of Finance*, 53(6), 2059-2106.

Engle, R. F., & Mustafa, C. (1992). Implied ARCH models from options prices. *Journal of Econometrics*, 52(1), 289-311.

Frijns, B., Lehnert, T., & Zwinkels, R. C. (2010). Behavioral heterogeneity in the option market. *Journal of Economic Dynamics and Control*, 34(11), 2273-2287.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2), 453-497.

Hao, J., & Zhang, J. E. (2013). GARCH Option Pricing Models, the CBOE VIX, and Variance Risk Premium. *Journal of Financial Econometrics*, 11(3), 556-580.

Härdle, W., & Hafner, C. M. (2000). Discrete time option pricing with flexible volatility estimation. *Finance and Stochastics*, 4(2), 189-207.

Hentschel, L. (1995). All in the family nesting symmetric and asymmetric garch models. *Journal of Financial Economics*, 39(1), 71-104.

Heston, S. L., & Nandi, S. (2000). A closed-form GARCH option valuation model. *Review of Financial Studies*, 13(3), 585-625.

Hsieh, K. C., & Ritchken, P. (2005). An empirical comparison of GARCH option pricing models. *Review of Derivatives Research*, 8(3), 129-150.

Hull, J., & Suo, W. (2002). A methodology for assessing model risk and its application to the implied volatility function model. *Journal of Financial and Quantitative Analysis*, 37(2), 297-318.

Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, 15(5), 850-859.

Lehar, A., Scheicher, M., & Schittenkopf, C. (2002). GARCH vs. stochastic volatility: Option pricing and risk management. *Journal of Banking & Finance*,26(2), 323-345.

Lehnert, T. (2003). Explaining smiles: GARCH option pricing with conditional leptokurtosis and skewness. *The Journal of Derivatives*, 10(3), 27-39.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, 23(2), 821-862.

Yatchew, A., & Härdle, W. (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap. *Journal of Econometrics*, 133(2), 579-599.

Table 1: Average Price, BS-Implied Volatility and number of contract by Moneyness and Maturity

			Maturity		
			<60	60 to 160	>60
Moneyness	<0.85	Price	0.769	2.534	8.597
		σ_{BS}	0.321	0.277	0.237
		Nb Contract	1766	2342	2800
	0.85-1	Price	8.389	19.063	38.584
		σ_{BS}	0.185	0.182	0.179
		Nb Contract	3341	2123	2273
	>1 to <1.15	Price	7.431	15.724	34.761
		σ_{BS}	0.145	0.142	0.143
		Nb Contract	2929	2111	2211
	>1.5	Price	0.339	0.840	3.900
		σ_{BS}	0.286	0.201	0.154
		Nb Contract	1571	2249	3048

Notes: the option sample characteristic averages are summarized by maturity and moneyness. Moneyness in this table is computed as (k/s). Since only out of the money options are conserved, all options with a moneyness lower than 1 are put options. All options with a moneyness higher than 1 are call options. Maturities are in day.σ_{BS} refers to the Black and Scholes implied volatility.

Table 2: Sample Composition and Market Conditions Descriptive Statistics

Panel A

	Mean	Std. Dev.	Min	Max
Market Conditions				
Past Returns	0.01	0.08	-0.23	0.25
Realized Volatility	0.17	0.09	0.07	0.54
VIX	0.21	0.07	0.11	0.42
Leverage	0.27	0.07	0.15	0.47
Local Volatility	0.17	0.06	0.08	0.36
Sample Compositions				
Average Price	11.53	2.03	6.80	16.78
Average Moneyness	0.19	0.04	0.13	0.35
Average Maturity	135.18	15.30	108.33	178.14
Call %	0.49	0.13	0.24	0.81
Cheap Option %	0.31	0.04	0.23	0.39

Panel B

	Past Returns	Realized Volatility	VIX	Leverage	Local Volatility	Average Price	Average Moneyness	Average Maturity	Call %	Cheap Option %
Past Returns	1									
Realized Volatility	-0.59***	1								
VIX	-0.64***	0.84***	1							
Leverage	0.07	-0.43***	-0.45***	1						
Local Volatility	-0.63***	0.83***	0.99***	-0.46***	1					
Average Price	-0.42***	0.54***	0.61***	-0.14*	0.62***	1				
Average Moneyness	-0.51***	0.75***	0.93***	-0.5***	0.94***	0.55***	1			
Average Maturity	0.04	0.1	0.07	-0.01	0.09	0.57***	0.22***	1		
Call %	-0.71***	0.74***	0.86***	-0.41***	0.86***	0.64***	0.78***	0.11	1	
Cheap Option %	-0.07	0.12	0.21***	-0.34***	0.21***	-0.42***	0.32***	-0.4***	0.18**	1

Notes: Standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

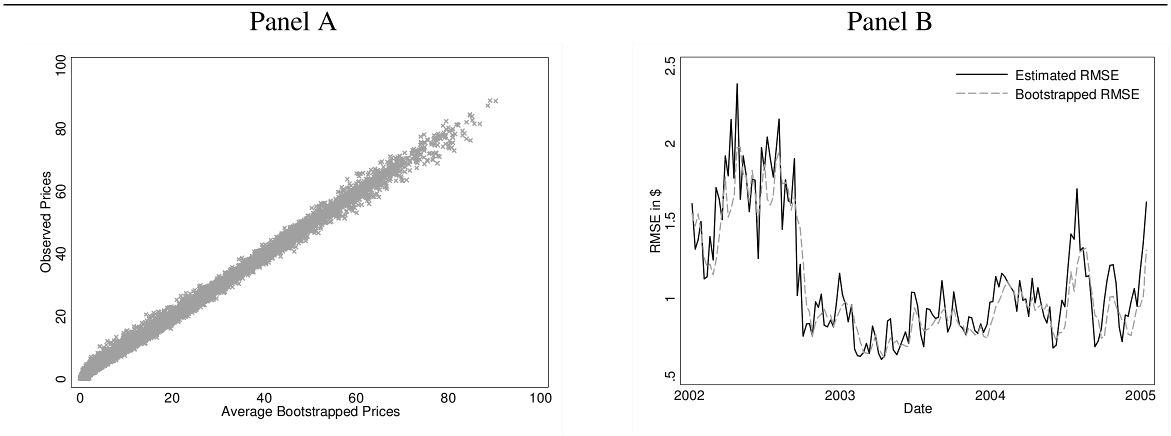
The table presents the descriptive statistics of Sample composition and market variables in Panel A. Past return and realized volatility are computed respectively over 3 months and 14 days. Past return and realized volatility are measured for each cross-section. Sample compositions refers to information about each of the 152 (2002 to 2004) samples used. Call % refers to the percentage of call option in the sample. Cheap Option % is the percentage of option below 1\$. Average moneyness is computed as the absolute value of (K/S)-1. Panel B reports the correlation matrix.

Table 3: Parameter Estimates and resulting in-sample RMSE

	α	β_0	β_1	Local Volatility	RMSE
Mean	-0.025	-0.288	0.272	16.5%	1.124
SD	0.01	0.065	0.066	6.4%	0.401
Min	-0.089	-0.569	0.149	8.2%	0.611
Max	-0.003	-0.155	0.466	35.7%	2.381
1st quartile	-0.029	-0.324	0.215	11.5%	0.832
3rd quartile	-0.02	-0.24	0.318	20.2%	1.327

Notes: The present statistics result from the 155 calibrations performed every Wednesdays. The statistic of all the parameters estimated and the loss function are displayed. The local volatility is expressed in yearly volatility.

Figure 1: Bootstrapped Prices and RMSEs



Note: Panel A plot the actual observed prices against the average of the bootstrapped prices for each observation. Panel B compare the time series of RMSE along different cross section. The estimated RMSE is obtained from the original estimation. The bootstrapped RMSE is the mean of the bootstrapped RMSE distribution.

Table 4: One sample T-test of the average Observed price-Average Bootrapped price difference per option category

		Maturity		
		<60	60 to 160	>60
Moneyness	<0.85	0.00	0.01	0.01
		(0.56)	*(0.67)*	*(0.16)*
	0.85-1	-0.01	-0.00	-0.01
		(0.67)	*(0.81)*	*(0.55)*
	>1 to <1.15	0.01	0.02	-0.01
		(0.30)	*(0.38)*	*(0.87)*
	>1.5	-0.01	-0.01	-0.01
		(0.21)	*(0.31)*	*(0.62)*

Note: The table present the average values of the difference between the observed prices and the average bootstrap equivalent prices. In brackets are the P-values associated with a 0 mean difference test.

Table 5: Time Series and Bootstrap parameters information comparison

Panel A: Standard Deviation					
	α	β_0	β_1	Local Volatility	RMSE
SD Time Series	0.01	0.065	0.066	0.064	0.401
SD Bootstrap	0.003	0.017	0.015	0.002	0.074

Notes: The 155 estimation performed are resulting in 155 parameters and loss functions at different point in time. The SD Time series is the standard deviation of these series.
The bootstrap enables to obtain a standard deviation for each estimation. Therefore 155 SD are obtained. The SD Bootstrap displayed is the average of these 155 SD.

Table 6: Correlation Matrix- Time Series and Cross Section

Time Series Correlations

	α	β_0	β_1	Local Volatility
α	1			
β_0	0.695***	1		
β_1	-0.504***	-0.857***	1	
Local Volatility	0.0826	0.0977	-0.456***	1

Notes: These are the Time Series Correlations of the 155 estimated parameters. The optimization is performed and a set of parameters is obtained every Wednesday. The correlations proposed are the correlation between all the parameters obtained over time, over multiple optimizations. Significance level are obtained with a Pearson Test

Cross-Section Correlations

	α	β_0	β_1	Local Volatility
α	1			
β_0	0.93***	1		
β_1	-0.941***	-0.996***	1	
Local Volatility	0.282***	0.228***	-0.27***	1

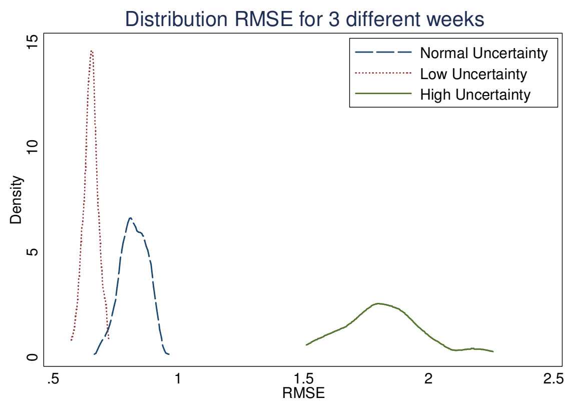
Notes: The values reported are the average of each cross section correlation. Each cross-sections are bootstrapped resulting in a set of parameters for each cross-section. Therefore, correlations are obtained at the cross-sectional level. This table present the average of all cross-sections correlations. The significance level is obtained with a t-test

		α	β_0	β_1	Local Volatility
α	Min	1			
	Max	1			
β_0	Min	0.784	1		
	Max	0.994	1		
β_1	Min	-0.99	-0.999	1	
	Max	-0.836	-0.984	1	
Local Volatility	Min	-0.83	-0.843	-0.913	1
	Max	0.937	0.868	0.807	1

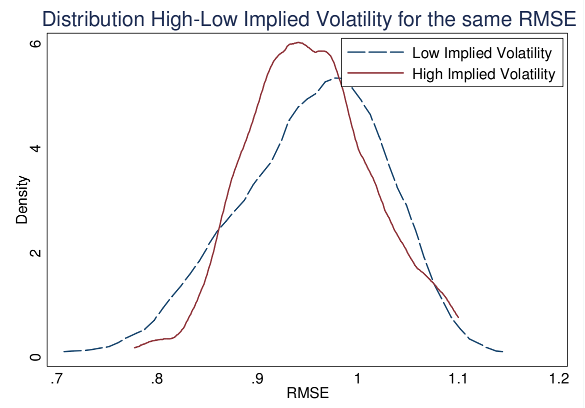
Notes: The minimum/maximum cross section correlation (obtained through bootstrap) are reported here

Figure 2: RMSE uncertainty represented as bootstrapped distribution in different contexts

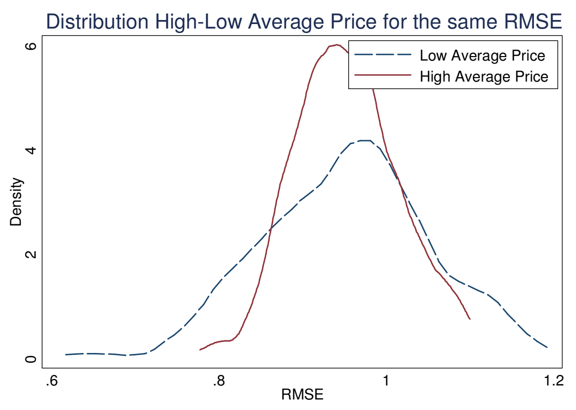
Panel A



Panel B



Panel C



Notes: Kernel density estimates are used to smooth the distribution.

Table 7: RMSE distributions in different context- Statistical test

	Mean Comparison	Variance Comparison	K-S
RMSE			
Normal-High	0.000	0.000	0.000
Normal-Low	0.000	0.000	0.000
High-Low	0.000	0.000	0.000
Implied Volatility			
High-Low	0.977	0.154	0.405
Average Option Price			
High-Low	0.963	0.000	0.12

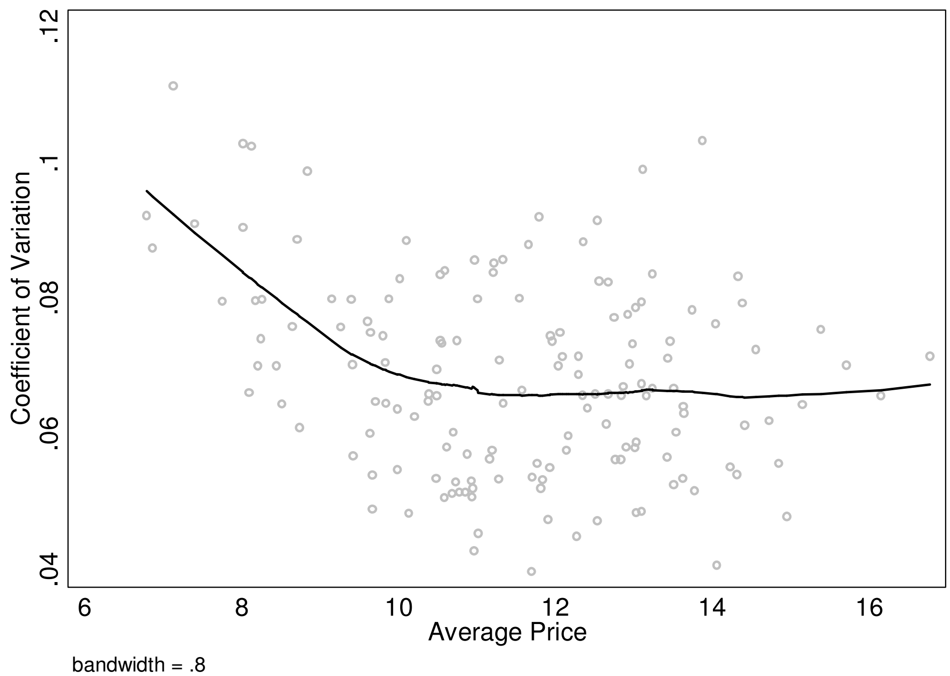
Notes: The table provides the P-value of alternative distribution comparison test (mean comparison test, variance comparison test, Kolomogrov-Smirnov test) for the previously displayed distributions.

Table 8: Cross Section Pricing Performance, Sample Composition and Market Conditions

	ASC			Δ RMSE	Coeficient of Variation		
	Autoregressive	Multivariate	Full		Autoregressive	Multivariate	Full
Δ Local Volatility		-0.289	-0.073	0.181***		0.021**	0.005
		(0.262)	(0.181)	(0.059)		(0.010)	(0.006)
Leverage		-1.173***	-0.719***	0.218**		0.060***	0.024**
		(0.420)	(0.268)	(0.102)		(0.017)	(0.011)
Average Price		-0.080***	-0.011	-0.012***		-0.002***	-0.001
		(0.014)	(0.012)	(0.004)		(0.001)	(0.000)
Cheap Option %		-1.423	-1.005*	-0.003		0.089**	0.045**
		(0.876)	(0.558)	(0.211)		(0.037)	(0.021)
Average Maturity		0.000	-0.004**	0.003***		0.000***	0.000***
		(0.002)	(0.001)	(0.001)		(0.000)	(0.000)
Dependant Variable Lag	0.533***		0.485***		0.839***		0.785***
	(0.088)		(0.087)		(0.047)		(0.046)
Dependant Variable Lag 2	0.275***		0.296***				
	(0.084)		(0.081)				
Constant	0.212***	2.760***	1.349***	-0.281**	0.011***	0.017	-0.016
	(0.070)	(0.527)	(0.322)	(0.117)	(0.003)	(0.021)	(0.011)
Observations	150	151	150	151	151	151	151
R-squared	0.578	0.192	0.615	0.224	0.683	0.196	0.707

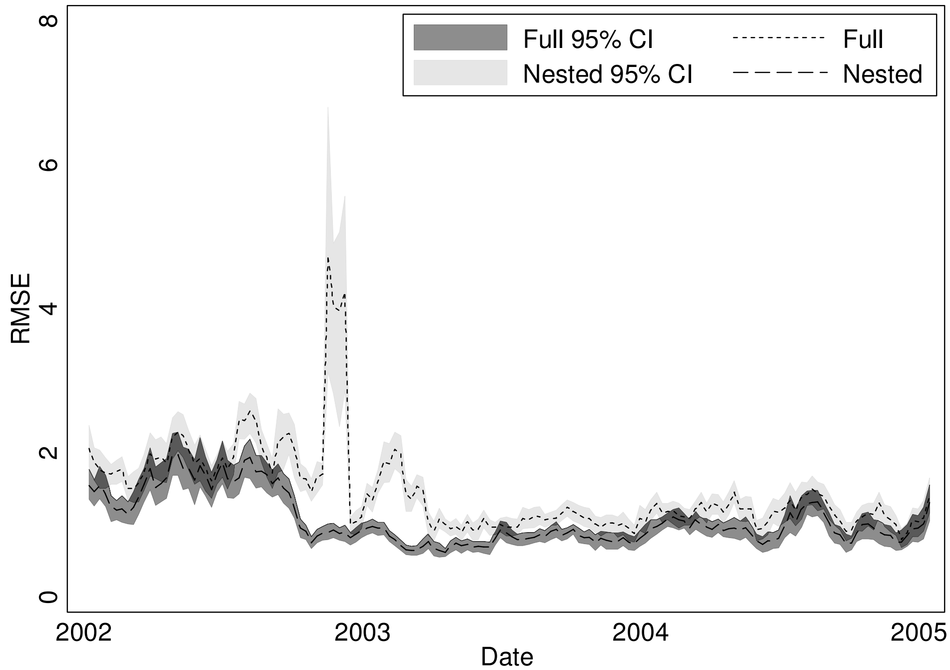
Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Local Volatility and RMSE appears to be non stationary based on a Dickey-Fuller test. As a precaution we take the first difference of these variables. 3 dependant variables are used ASC, Δ RMSE and Coefficient of Variation. For ASC and Coefficient of variation 3 models are performed: a simple auto regressive model, a multivariate model and a combined model. For Δ RMSE only one model is available because as a first difference this time series is not autoregressive. For ASC and Coefficient of Variation the lag length is chosen based on previous time series analysis not reported here.

Figure 3: Very inexpensive options and uncertainty



Notes: The figure presents the relationship between the coefficient of variation and Average price. The locally weighted regression of coefficient of variation on the average price line is also displayed.

Figure 4: Cross-sectional specification tests



Notes: This figure is the graphic representation of the specification test. The light grey interval represents the 95% confidence interval of the nested model. The darker grey interval represents the 95% confidence interval of the full model. The darkest areas indicate where the two RMSE distributions are overlapping. A sufficient overlap is consistent with a statistical rejection of distinguishing the two models' RMSE.