

RISK PREMIUM EFFECTS ON IMPLIED VOLATILITY REGRESSIONS

Leonidas S. Rompolis

Department of Public and Business Administration
University of Cyprus

Elias Tzavalis

Department of Economics.
Athens University of Economics and Business

Abstract

This article provides new insights into the sources of bias of option implied volatility to forecast its physical counterpart. We argue that this bias can be attributed to volatility risk premium effects. The latter are found to depend on high-order cumulants of the risk-neutral density. These cumulants capture the risk-averse behavior of investors in the stock and option markets for bearing the investment risk that is reflected in the deviations of the implied risk-neutral distribution from the normal distribution. We show that the bias of implied volatility to forecast its corresponding physical measure can be eliminated when the implied volatility regressions are adjusted for risk premium effects. The latter are captured mainly by the third-order risk-neutral cumulant. We also show that a substantial reduction of higher-order risk-neutral cumulants biases to predict their corresponding physical cumulants is supported when adjustments for risk premium effects are made.

JEL Classification: C10, G13, G14.

Note: The authors would like to thank an anonymous referee and the editor of this journal for their very useful comments and suggested extensions on an earlier version of the paper. Leonidas S. Rompolis acknowledges financial support from the Iraklitos Fellowships of Research of the Athens University of Economics and Business.

1. Introduction

There is voluminous evidence that the volatility implied by option prices, referred to as risk-neutral volatility, constitutes a biased upward predictor of the future realized volatility of asset (stock) returns, known as physical volatility, over the remaining life of the relevant options (see, e.g., Canina and Figlewski 1993, Poon and Granger 2003, and Jiang and Tian 2005). This bias means that the slope coefficient of a linear regression of physical volatility on implied volatility will be biased downward from unity. This evidence is in contrast to the pure (without risk premium) version of the market efficiency hypothesis asserting that the preceding two measures of volatility should possess a one-to-one correspondence. As is shown in many recent studies, the implied volatility bias can be related to the market volatility risk premium (see Bates 2003, Bollerslev and Zhou 2006), creating a wedge between physical volatility and risk-neutral volatility. In particular, based on the stochastic volatility option pricing model, Bollerslev and Zhou (2006) show that the bias of the slope coefficient of the standard implied volatility regression is not zero and critically depends on the parameters of the underlying option pricing model.

The present article provides new insights into the sources of the volatility risk premium and the forecasting bias of the slope of the implied volatility regressions. We derive a theoretical relation between physical volatility and risk-neutral volatility, which shows that under a power utility pricing kernel, the difference between the two measures of volatility depends on higher than second-order risk-neutral cumulants. This relation is derived in a model-free manner, without imposing any structure on the underlying asset price process or adopting any specific parametric option pricing model (see Bakshi and Madan 2000). It can also be easily extended to analogous relations predicting the future level of higher-order cumulants based on their risk-neutral counterparts.

Expressing the difference between physical and risk-neutral volatilities in terms of cumulants rather than moments enables us to distinguish its generating sources more clearly. In particular, the cumulants greater than second-order quantify deviations from the normal distribution and therefore measure large risks beyond the variance (equal to the second-order cumulant). The third and higher odd-order cumulants capture asymmetric deviations, and the fourth and higher even-order cumulants reflect symmetric deviations. Because cumulants generally constitute adequate measures of risk (see Malevergne and Sornette 2006), the higher-order risk-neutral cumulants that enter the physical-risk-neutral volatility relation derived in this article can be given the interpretation of reflecting risk premium effects. These compensate investors in the asset markets for bearing the risk of deviations from the normal distribution of asset returns due, for instance, to asymmetric or extreme values of them. Among these effects, we provide strong evidence that the risk premium captured by the third-order risk-neutral cumulant, which is associated with the skewness of the risk-neutral density, can sufficiently explain the bias of the slope coefficient of the standard implied volatility regression. This is shown in a new regression framework adjusted for the risk premium effects and can be done for small estimates of the relative risk aversion coefficient. This framework can be easily extended to test for the information content of higher than second-order risk-neutral cumulants about their physical counterparts.

II. On the Relation Between Risk-Neutral and Physical Volatility

Theoretical Model

Consider an economy in the absence of arbitrage. Then, there will be a pricing kernel projected onto the asset (stock) log-return, denoted as $\rho_T = \ln\left(\frac{S_T}{S_t}\right)$ where S_T (or S_t) denotes the asset price at the future time T (or at the current time t). Under this kernel, the physical probability density function of ρ_T conditional on time t information set (denoted as L_t) is related to its risk-neutral counterpart through the following relation:

$$p_t(\rho_T) = \frac{q_t(\rho_T) / M_t(\rho_T)}{\int q_t(\rho_T) / M_t(\rho_T) d\rho_T}, \quad (1)$$

where $p_t(\rho_T)$ and $q_t(\rho_T)$ denote the conditional physical and risk-neutral probability densities of ρ_T and $M_t(\rho_T)$ is the projected pricing kernel onto the payoff ρ_T . Under the power/logarithmic utility, $M_t(\rho_T)$ is given as $M_t(\rho_T) = e^{-\gamma\rho_T}$, where γ is the relative risk aversion coefficient. Then, relation (1) can be written as:

$$p_t(\rho_T) = \frac{e^{\gamma\rho_T} q_t(\rho_T)}{\int e^{\gamma\rho_T} q_t(\rho_T) d\rho_T} \quad (2)$$

where the normalizing factor $\int e^{\gamma\rho_T} q_t(\rho_T) d\rho_T$ integrates the values of $p_t(\rho_T)$ to unity. Based on (2), in the next proposition we derive a theoretical relation linking the v^{th} -order physical cumulant of the log-return ρ_T conditional on the information set L_t to the risk-neutral cumulants¹.

Proposition 1 Let $k_{\rho,n}^N$ and $k_{\rho,n}^P$ be the v^{th} -order cumulants of the conditional on the current market information set L_t risk-neutral probability density function $q_t(\rho_T)$ and the physical probability density function $p_t(\rho_T)$ of the log-return ρ_T , respectively. Then, the following relation holds:

$$k_{\rho,n}^P = \sum_{m=0}^{\infty} \left(\frac{\gamma^m}{m!} \right) k_{\rho,n+m}^N, \quad \forall v \in \mathbb{N} \quad (3)$$

where $\frac{\gamma^0}{0!} = 1$.

For $v = 2$, the theoretical relation given by Proposition 1 implies that the second-order cumulants $k_{\rho,2}^P$ and $k_{\rho,2}^N$, which respectively represent physical volatility and risk-neutral volatility, are linked through the following relation:

$$k_{\rho,2}^P = k_{\rho,2}^N + \gamma k_{\rho,3}^N + \frac{\gamma^2}{2!} k_{\rho,4}^N + \frac{\gamma^3}{3!} k_{\rho,5}^N + \dots \quad (4)$$

implying that

$$k_{\rho,2}^P - k_{\rho,2}^N = \gamma k_{\rho,3}^N + \frac{\gamma^2}{2!} k_{\rho,4}^N + \frac{\gamma^3}{3!} k_{\rho,5}^N + \dots \quad (5)$$

From the last relation, it can be easily seen that there is no one-to-one correspondence between risk-neutral and physical volatilities captured by the second-order cumulants $k_{\rho,2}^P$ and $k_{\rho,2}^N$, respectively. The two volatilities differ from each other, and their difference $k_{\rho,2}^P - k_{\rho,2}^N$, referred to as the volatility risk premium, critically depends on all higher than second-order cumulants of the risk-neutral density of the log-return ρ_T . Only when $\gamma = 0$ (risk-neutrality) do the values of $k_{\rho,2}^P$ correspond to those of $k_{\rho,2}^N$ in a one-to-one

¹ Note that the first four cumulants, k_n for $n = 1, 2, 3, 4$ of a probability density function are related to the central (μ'_m) and non-central (μ_n) moments through the following relationships: $k_1 = \mu'_1 = \mu_1$, $k_2 = \mu'_2 - (\mu_1^2)$, $k_3 = \mu'_3 - (3\mu_2\mu_1 - 2\mu_1^3)$ and $k_4 = \mu'_4 - 3\mu_2^2 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 2\mu_1^4$, (see Stuart and Ord 1994). Based on the above definitions of cumulants, we can write the skewness and kurtosis coefficients as $sk = \frac{k_3}{(k_2)^{3/2}}$ and $Ku = 3 + \frac{k_4}{(k_2)^2}$.

relationship. The latter can also take place when $k_{\rho,2}^N = 0$, for all $\mu > 2$, which holds under the assumptions of the Black and Scholes (BS) model.

In contrast to the moments, the cumulants of the risk neutral-density that feature in the right-hand side of relation (5) can directly measure deviations of this distribution from the normal one, with the third and higher odd-order cumulants to capturing asymmetric deviations, and the fourth and higher even-order cumulants to reflecting symmetric deviations. In fact, this relation enables us to separate the implied volatility risk premium effects caused by the above distinct sources of deviations from the Gaussian law: skewness and excess kurtosis. By their definitions (see footnote 3), the former is directly related to the third-order cumulant, and the latter to the fourth-order cumulant.

Risk-neutral skewness commands a volatility risk premium to compensate risk-averse investors for bearing the risk associated with asymmetric stock return distributions. A negative value of risk-neutral skewness often found in practice (implying $k_{\rho,3}^N < 0$) should be taken to reflect the fact that risk-averse investors assign higher risk-neutral probabilities to negative stock returns (or, more generally, to stock returns with smaller values than their mean) than positive returns. In forecasting expected stock market physical volatility $k_{\rho,2}^P$, the previously described behavior of risk-averse investors will have the following consequences. To estimate $k_{\rho,2}^P$, an upward adjustment must be made to the risk-neutral probabilities of positive stock returns (which are smaller than the risk-neutral probabilities of negative returns) and a downward adjustment to the risk-neutral probabilities of negative returns (which are larger to the risk-neutral probabilities of positive returns). Due to the exponential nature of the pricing kernel implied by the power utility, the upward adjustment is approximately equal to the downward adjustment around the mean of the risk-neutral density and both of them are proportional to the risk aversion coefficient ϕ .

Thus, if risk-neutral skewness is negative, expected physical volatility will be smaller than risk-neutral volatility by a factor which equals ϕ -times the risk-neutral third-order cumulant, $k_{\rho,3}^N$, as is predicted by relation (5). In this case, the volatility risk premium $k_{\rho,2}^P - k_{\rho,2}^N$ will be negative. Note that the higher the risk aversion coefficient ϕ , the higher the negative effect of $k_{\rho,3}^N$ on $k_{\rho,2}^P - k_{\rho,2}^N$ will be.

By analogous arguments as above, we can explain the effect of the fourth-order riskneutral cumulant $k_{\rho,4}^N$ on $k_{\rho,2}^P - k_{\rho,2}^N$. This cumulant reflects a separate source of volatility risk premium. A positive value of $k_{\rho,4}^N$, implying that the risk-neutral density has fat tails (or excess kurtosis), can be attributed to the fact that risk-averse investors assign higher probabilities to extreme values of stock returns². Thus, in calculating physical volatility an upward adjustment should be made to the risk-neutral probabilities of positive extreme values of stock returns distributions and a downward adjustment to the risk-neutral probabilities of negative returns. Since these probabilities concern stock returns whose values are not close to their mean, the exponential nature of the pricing kernel of the power utility function (see relation (2)) implies that the upward adjustment of the risk-neutral probabilities will be higher than the downward one. This can explain the positive relationship between the volatility risk premium $k_{\rho,2}^P - k_{\rho,2}^N$ and the fourth-order risk-neutral cumulant $k_{\rho,4}^N$, implied by relation (5). Our theoretical relation (3) can be also employed to investigate the sources of the risk premium reflected into the difference between higher than second-order physical and riskneutral cumulants, i.e. $k_{\rho,n}^P - k_{\rho,n}^N$. In particular, for the third and fourth-order cumulants, which are related to skewness and excess kurtosis, respectively, relation (3) implies:

$$k_{\rho,n}^P - k_{\rho,n}^N = \gamma k_{\rho,n+1}^N + \frac{\gamma^2}{2!} k_{\rho,n+2}^N + \frac{\gamma^3}{3!} k_{\rho,n+3}^N + \dots, \text{ for } v = 3, 4. \quad (6)$$

The last relation is analogous to (5). It shows that the existence of the third-order cumulant risk premium, $k_{\rho,3}^P - k_{\rho,3}^N$, also referred to as the skewness premium, can be explained by higher than third-order risk-neutral cumulants³. According to this, a positive value of the

² The economic intuition of using the fourth-order cumulant as a measure of risk is given by Malevergne and Sornette (2006). In the context of the risk-neutral density, this can be obtained by writing the fourthorder cumulant $k_{\rho,4}^N$ as $k_{\rho,4}^N = \mu_{\rho,4}^N - 3(\mu_{\rho,2}^N)^2$, where $\mu_{\rho,4}^N$ and $\mu_{\rho,2}^N$ are the fourth and second-order risk-neutral central moments, respectively. A positive value of $k_{\rho,4}^N$ implies the following inequality $\mu_{\rho,4}^N > 3\mu_{\rho,2}^N$. This can be interpreted as expressing the fact that investors are more risk-averse to large risks associated with extreme values of stock returns measured by $\mu_{\rho,4}^N$, compared to smaller ones related to smaller variation of stock returns measured by $\mu_{\rho,2}^N$. Having $k_{\rho,4}^N = 0$ (implied by the BS model) means that risk-averse investors are indifferent between large and small risks.

³ Bakshi, Kapadia and Madan (2003) derived an approximating version of this relation based on the risk-neutral skewness and kurtosis coefficients.

skewness premium (implying that the physical distribution is more skewed towards the right compared to the risk-neutral distribution) can be attributed to the positive sign of the fourth-order risk-neutral cumulant $k_{\rho,4}^N$ capturing the risk-averse behavior of investors to assign higher probabilities to extreme values of stock returns, as mentioned before. Analogously, we can explain a negative sign of the excess kurtosis premium reflected into the difference $k_{\rho,4}^P - k_{\rho,4}^N$.

Explaining the Bias of the Standard Implied Volatility Regressions

The results of Proposition 1 can be used to investigate the sources of documented biases in forecasting future levels of physical volatility based on currently observed levels of risk-neutral volatility. Relation (5) provides us with the theoretical underpinnings of the volatility risk premium and, thus, permits to specify a new econometric framework to measure the slope bias of the implied volatility regressions. This framework only requires estimates of the physical and risk-neutral cumulants of the log-return ρ_T , which can be easily obtained from the data. As mentioned before, the statistical framework which is often employed in practice to forecast the future level of physical volatility based on implied volatility is a single regression model employing the risk-neutral volatility $k_{\rho,2}^N$ (say with a maturity interval $T-\tau$) as a regressor and the future level of physical volatility $k_{\rho,2}^P$, over the future period $T-\tau$, as a regressand, i.e.

$$k_{\rho,2}^P = \alpha + \beta \cdot k_{\rho,2}^N + e, \quad (7)$$

where e stands for a zero mean forecast error conditional on the current market information set L_t . Estimation of this regression model can be performed using the least squares method with time series data on the second-order cumulants $k_{\rho,2}^N$ and $k_{\rho,2}^P$. These estimates can then be used to test for the validity of the following two hypotheses. First, whether implied volatility contains some information about the future value of physical volatility, which is often measured by the future realized volatility. In this case, the estimate of the slope coefficient β should be different from zero. Second, whether implied volatility constitutes an unbiased forecast of physical volatility, which implies the following joint hypothesis: $\alpha = 0$ and $\beta = 1$. Under the last hypothesis, deviations of the regression estimates of coefficients α and β from their theoretical values $\alpha = 0$ and $\beta = 1$ are considered as evidence of bias and information inefficiency in option market forecasts. However, in the light of relation (5), one ought to interpret such a failure of option market forecasts with caution. A rejection of the joint hypothesis $\alpha = 0$ and $\beta = 1$ (or the hypothesis $\beta = 1$) can be attributed to the risk premium effects omitted from the right-hand side of the implied volatility regression model (7). As equation (5) shows, the risk-neutral volatility $k_{\rho,2}^N$ will fully reflect information about its physical counterpart only when $\phi = 0$ (i.e. under risk neutrality) or when the BS model holds, implying that $k_{\rho,n}^N = 0$ for all $n > 2$. In contrast, if $\phi \neq 0$ and $k_{\rho,n}^N \neq 0$, for $n > 2$, then the least squares estimator of the slope coefficient β will be biased, while that of the intercept α may give a value different from zero⁴.

The bias of the least squares estimator of the slope coefficient β will depend on the magnitude of the risk aversion coefficient ϕ , the sign and size of the correlation coefficients between the second-order risk-neutral cumulant $k_{\rho,2}^N$ and the higher-order cumulants $k_{\rho,2+m}^N$, for $\mu=1,2,\dots$, which are omitted from the regression model (7). Moreover, the bias will also depend on the standard deviation of $k_{\rho,2+m}^N$ relative to that of $k_{\rho,2}^N$. Below, we present an analytic formula of the asymptotic bias of the estimator of β , denoted as $\hat{\beta}$, from unity:

$$AsyBias(\hat{\beta}) \equiv p \lim_{T \rightarrow \infty} \hat{\beta} - 1 = \sum_{m=0}^{\infty} \left(\frac{\gamma^m}{m!} \right) \phi_{2,2+m} \frac{SD(k_{\rho,2+m}^N)}{SD(k_{\rho,2}^N)} \hat{\beta} \quad (8)$$

where T denotes here the number of the time series observations of the sample, $\phi_{2,2+m}$ stands for the correlation coefficient between the risk-neutral cumulants $k_{\rho,2}^N$ and $k_{\rho,2+m}^N$, for all $\mu \in 1$, and $SD(\cdot)$ stands for standard deviation. Formula (8) implies that, under risk aversion (i.e. $\phi \neq 0$), estimator $\hat{\beta}$ will be biased downward from unity if some of the correlation coefficients $\phi_{2,2+m}$ are negative. This source of bias can explain the upward bias of the implied volatility to forecast the future level of its physical counterpart documented in

⁴ This will happen due to the omission of higher than second-order risk-neutral cumulants from the righthand side of (7). Alternative estimators, such as the instrumental variables (IV) or the generalized methods of moments (GMM) suggested in the literature to deal with the measurement errors in the risk-neutral volatility (see Christensen and Prabhala 1998) can not resolve this problem, as most of the instruments used for this purpose are correlated with the omitted risk-neutral cumulants from (7).

the literature. Note that, only when $\lambda_{2,2+\mu} = 0$ or $\frac{SD(k_{\rho,2+m}^N)}{SD(k_{\rho,2}^N)}$ is close to zero, $\hat{\beta}$ will provide unbiased estimates of the slope

coefficient β of regression (7) under risk-averse behavior. To calculate the magnitude of the asymptotic bias of the estimator $\hat{\beta}$ and, hence, measure the risk premium effects on β , formula (8) shows that an unbiased estimator of the risk aversion coefficient α is required.

III. Empirical Analysis

The main goal of this section is to assess the ability of our theoretical relation (4) to explain the documented biases of implied volatility regressions to forecast future levels of physical volatility. To this end, we suggest a new implied volatility regression model allowing for risk premium effects. This model can be also used to provide estimates of the risk aversion coefficient implied by option and stock markets data. Our analysis is also extended to investigate whether relation (6) can consistently explain the biases of higher than second-order risk-neutral cumulants to forecast their corresponding physical cumulants.

Data

Our empirical analysis uses index returns and cross-section sets of monthly observations on option prices for three different indexes: Standard & Poor's 500 Index (henceforth denoted as S&P500) and Financial Times Stock Exchange 100 Index (henceforth denoted as FTSE100) from 1996:01 to 2007:12, as well as NASDAQ 100 Index (henceforth denoted as NASDAQ) from 2001:06 to 2007:12. These sets of option prices cover different strike prices and have maturity intervals of approximately one calendar month (i.e. twenty two trading days). By employing these data, we have obtained estimates of the risk-neutral cumulants $k_{\rho,n}^N$ at any point in time τ of our sample based on the formulas suggested by Bakshi, Kapadia and Madan (2003), and Rompolis and Tzavalis (2008). These formulas give estimates of non-central risk-neutral moments from out-of-the-money option prices in a model-free manner. Employing them we have derived values of the risk-neutral cumulants using Stuart's and Ord (1994) cumulant formulas (see footnote 3). As shown by Rompolis and Tzavalis (2008), this procedure of retrieving the risk-neutral cumulants $k_{\rho,n}^N$ from option prices provides very precise estimates of them⁵.

As values of the physical volatility $k_{\rho,2}^P$ or the higher-order cumulants $k_{\rho,3}^P$ and $k_{\rho,4}^P$, our analysis uses unbiased cumulant estimates of them based on the future realized values of the three indexes daily returns, denoted as ρ_j for $j = 1, 2, \dots, n_T$, up to one-month head, i.e. for the time interval of $n_T = 22$ approximately trading days for each month (see, e.g., Canina and Figlewski 1993). For instance,

the realized volatility is calculated as $k_{\rho,2}^P = \frac{1}{n_T - 1} \sum_{j=1}^{n_T} (p_j - \bar{p})^2$, where $\bar{p} = \frac{1}{n_T} \sum_{j=1}^{n_T} \rho_j$. Analogous relations to the above can be

employed to estimate the third and fourth-order physical cumulants (see Stuart and Ord 1994). The estimates of physical cumulants derived by the above method have the key feature that they result in nonoverlapping data, as time periods covered by successive options exhibit no overlap. The nonoverlapping nature of the data can yield more precise estimates of them relative to their historical estimates obtained from overlapping samples (see Christensen and Prabhala 1998).

Estimation of Implied Volatility and Higher-Order Risk-Neutral Cumulant Regressions With and Without Risk Premium Effects

In this section we introduce a new implied volatility regression framework adjusted for risk premium effects which provides unbiased forecasts of physical volatility. This framework can be also extended to examine the information content of higher than second-order riskneutral cumulants.

Implied Volatility Regressions Without Risk Premium Effects. We commence our analysis with presenting estimates of the standard implied volatility regression model (7), without adjusting for risk premium effects. We also present estimates of other two implied volatility regression models often used in practice to forecast functions of the realized volatility $k_{\rho,2}^P$. These models employ the square root

transformation of $k_{\rho,2}^P$ and $k_{\rho,2}^N$ (i.e. $(k_{\rho,2}^P)^{1/2}$ and $(k_{\rho,2}^N)^{1/2}$, or alternatively their logarithmic values $\ln[(k_{\rho,2}^P)^{1/2}]$ and $\ln[(k_{\rho,2}^N)^{1/2}]$ as regression variables (see, e.g., Andersen et al 2001, Corrado and Miller 2005, Giot and Laurent 2007). Under the risk-

⁵ For instance, the authors show that the percentage error of the risk-neutral kurtosis coefficient is, on average, less than 1.5%. This error is even smaller for the risk-neutral skewness coefficient.

neutrality hypothesis (i.e. $\alpha = 0$), note that the last two implied volatility regression models mentioned above can be thought of as direct transformations of our theoretical relation (5), underlying regression model (7).

Least squares (LS) estimates of all implied volatility regression models specified above are reported in Table 1. In particular, the table reports estimates of the coefficients α and β , the Wald test statistic of the joint hypothesis $\alpha = 0$ and $\beta = 1$, denoted as $W(\alpha = 0, \beta = 1)$, the adjusted coefficient of determination $\bar{R}^2$ and the Durbin-Watson ($\Delta\Omega$) statistic, as well as of two measures assessing the in- and out-of-sample forecasting performance of the models: the *RMSE* (root mean square error) and the *MAE* (mean absolute error). These measures are denoted as $RMSE^{(i)}$ and $MAE^{(i)}$ for the in-sample forecasts and as $RMSE^{(o)}$ and $MAE^{(o)}$ for the out-of-sample forecasts. They can be used to assess which one of the three implied volatility regression models considered in our analysis can provide better fit of our data, given that the coefficient $\bar{R}^2$ can not be used for this purpose when regression models are not nested (see Davidson and MacKinnon 1993). To examine this, the table also reports the values of the P_E statistic testing whether the standard regression model (7) or that based on the square-root variables $(k_{\rho,2}^P)^{1/2}$ and $(k_{\rho,2}^N)^{1/2}$ constitute better specifications of the data than the log-linear model. To calculate the out-of-sample forecasting performance measures $RMSE^{(o)}$ and $MAE^{(o)}$, we have estimated the alternative implied volatility models recursively starting from date 2004:01 and adding one observation at a time until the end of the sample. The out-of-sample forecasts are then calculated as the one-period ahead predictions of the models at each point of the recursive estimation.

The results of Table 1 are consistent with those provided in the literature. They indicate that, although the slope coefficient β of the standard implied volatility regression model (7) is of the right direction, its estimate is substantially less than unity, thus implying that the risk-neutral cumulant $k_{\rho,2}^N$ is a biased upward predictor of the future level of its physical counterpart $k_{\rho,2}^P$. Furthermore, the values of the Wald test statistic $W(\alpha = 0, \beta = 1)$ reported in the table clearly reject the joint hypothesis $\alpha = 0$ and $\beta = 1$ at 5% significance level. These results are stronger for the S&P500 and FTSE100 indexes. However, for NASDAQ the joint hypothesis $\alpha = 0$ and $\beta = 1$ is marginally rejected at the 5% significance level. Regarding the two other functional forms of the implied volatility regressions examined, the results of the table indicate that the estimates of the slope coefficient β are closer to unity, especially for the log-linear model. However, the estimates of these two models reject the joint hypothesis $\alpha = 0$ and $\beta = 1$. They also indicate that the forecasting performance of these two models is poorer than that of the standard implied volatility regression model, given by relation (7). Note that, among the three models, the highest prediction errors are observed for the log-linear model. Further evidence that this specification of the implied volatility regression model does not constitute a better specification of the data than the standard model is provided by the values of the P_E test statistic reported in the table. These can not reject the null hypothesis that the standard implied volatility regression model (or its square transformation) constitutes a correct specification of the data series, when compared to the log-linear model. One explanation that may be given to the above results and, in particular, to the evidence that the slope coefficient β tends to unity for the log-linear model is that, by taking the logarithmic transformation of the regression series, the risk premium effects are smoothed out. Of course, this is not without cost. As our results show, the estimates of the constant of the log-linear model become significantly different than zero and take negative values. The latter may explain the less than satisfactory performance of the log-linear model of implied volatility to forecast future levels of physical volatility, as demonstrated by the PMSE and MAE measures.

Finally, before closing this section we examine whether higher than second-order riskneutral cumulants provide unbiased predictions of their physical counterparts. To this end, we have estimated the following regressions:

$$k_{\rho,n}^P = \alpha + \beta k_{\rho,n}^N + u_n, \text{ for } n = 3, 4. \quad (9)$$

Least squares estimates of these regressions are reported in Table 3. These show that the bias of the slope coefficient β is worse than in the case of the standard implied volatility regression model (7). The estimates of the slope coefficient β of these models are not different than zero for all indexes considered, and thus the joint hypothesis $\alpha = 0$ and $\beta = 1$ is clearly rejected at 5% significance level. These results clearly indicate that the higherorder cumulants $k_{\rho,3}^N$ and $k_{\rho,4}^N$ contain no information about their physical counterparts. In the light of our theoretical relation (3) and the results of Table I (see Panels A and B), these biased estimates of β can be explained by the much higher correlation which exists among the higher-order risk-neutral cumulants $k_{\rho,3}^N$, $k_{\rho,4}^N$ and $k_{\rho,5}^N$, as well as

their relative 13 levels of standard deviations $\frac{SD(k_{\rho,4}^N)}{SD(k_{\rho,3}^N)}$ and $\frac{SD(k_{\rho,5}^N)}{SD(k_{\rho,4}^N)}$.

Implied volatility regressions allowing for risk premium effects. To investigate whether the risk premium effects can remove the bias of implied volatility regressions to forecast physical volatility, in this section we suggest a new statistical framework which, in light of our theoretical relation (5), includes higher-order risk-neutral cumulants in the standard regression model (7) as additional regressors. In particular, we have estimated the following regression model:

$$k_{\rho,2}^P = \alpha + \beta k_{\rho,2}^N + \gamma k_{\rho,3}^N + \varepsilon_2 \quad (10)$$

This model assumes that the third-order risk-neutral cumulant $k_{\rho,3}^N$, which captures the skewness effects on volatility premium, can account for the risk premium effects on the slope coefficient of the risk-neutral cumulant $k_{\rho,2}^P$. Note that, in addition to $k_{\rho,3}^N$, higher than second-order risk-neutral cumulants (e.g. $k_{\rho,3}^N$, $k_{\rho,4}^N$...) can be also included in the standard regression model (7), but as our initial estimates have shown the results of our analysis do not change. The inclusion of higher-order risk-neutral cumulants in regression (10) does not provide a better fit for our data. As mentioned before, estimation of the multiple regression model (10) should be performed with caution due to the multicollinearity problem that may appear in practice. This is due to the high correlation which exists between the two regressors $k_{\rho,2}^N$ and $k_{\rho,3}^N$. When this problem arises, the estimates of the slope coefficients of model (10) and/or their standard errors become very large (see Greene 2004). To see if this problem is present in our case, we have also estimated a restricted version of model (10) assuming that $\beta = 1$, i.e.

$$k_{\rho,2}^P - k_{\rho,2}^N = \alpha + \gamma k_{\rho,3}^N + \varepsilon_2^* \quad (11)$$

Comparison of the least squares estimates of coefficient γ between the above two models (10) and (11) can reveal whether any serious problem of multicollinearity is present in the unrestricted estimation of model (10). If this happens, then the estimates of γ should considerably differ between models (10) and (11).

The estimation results of models (10) and (11) are reported in Table 2. As can be seen from them, the inclusion of the third-order cumulant $k_{\rho,3}^N$ in the implied volatility regression model removes the bias of the least squares estimates of the slope coefficient β . The estimates of this coefficient now become close to unity and the joint hypothesis $\alpha = 0$ and $\beta = 1$ can no longer be rejected by the Wald test statistic $W(\alpha = 0, \beta = 1)$ at the 5% significance level. This is true for all indexes examined. Further support for the validity of hypothesis $\beta = 1$ and model (10) can be obtained by the values of the likelihood test statistic, ΔP , reported in the table. This test examines the validity of the restricted model (11), under the restriction $\beta = 1$, against its unrestricted version given by equation (10). As the results of the table show, this statistic can not reject the restricted model.

Another interesting result of Table 2 concerns the small values of the estimates of the risk aversion coefficient γ implied by our data sets. These values are very close for both the restricted and the unrestricted least squares estimates of models (11) and (10), respectively. They are also consistent with those found in other studies in the literature based on stock and/or option price data (see, e.g., Hansen and Singleton 1984, and Bliss and Panigirtzoglou 2004). They suggest that the risk aversion coefficient of the representative agent implied by option prices is not as extreme, as those found in studies based on consumption data (see, e.g. Mehra and Prescott 1985). The fact that the estimates of γ remain almost the same across the restricted and unrestricted regression models (10) and (11) can be taken as evidence of no presence of multicollinearity in the estimates of the coefficients β and γ of model (10).

To examine if the above estimates of γ are also consistent with those implied by our data for higher than second-order risk-neutral cumulants, in Table 3 we present least squares estimates of the restricted model (11) estimated as a system of the following equations:

$$k_{\rho,n}^P - k_{\rho,n}^N = \alpha + \gamma k_{\rho,n+1}^N + e_n, \quad n = 2, 3, 4. \quad (12)$$

These estimates are then used to calculate the values of the Wald statistic testing the hypothesis that the risk aversion coefficients γ , denoted as γ_n for each equation of system (12), are equal per pair of equations (i.e. $\gamma_2 = \gamma_3$ or $\gamma_3 = \gamma_4$) or across all of them (i.e. $\gamma_2 = \gamma_3 = \gamma_4$). The results of the table clearly indicate that the estimates of γ are very close for all equations of system (12) and the three stock indexes considered. The estimates of γ vary between the values 1.78 and 3.15, which are very close to those found in the augmented implied volatility regression (10). These results can be more formally confirmed by the Wald test statistics reported in the table which can not reject the above hypotheses examined. The only exemption is the joint hypothesis $\gamma_2 = \gamma_3 = \gamma_4$ for the FTSE100 index, which is rejected at the 5% significance level. But, even for this case the estimates of γ are very close to each other for all equations of the system. Note also that for this index, the pairs of hypotheses $\gamma_2 = \gamma_3$ and $\gamma_3 = \gamma_4$ can not be rejected when they are tested, separately.

Summing up, the results of this subsection clearly indicate that a relatively small degree of risk aversion in the options and stock markets can explain the biases of implied volatility to forecast its physical counterpart documented in the literature.

IV. Conclusions

In this article we provide a new framework of explaining the sources of bias of the implied by option prices volatility regressions to forecast future levels of physical volatility. This is done in a model-free manner, without imposing any structure on the stochastic process driving asset prices, or relying on any specific parametric option pricing model. The paper 17 derives a theoretical relation between physical and risk-neutral volatilities which shows that, under risk aversion, these two different measures of market volatility do not possess one-to-one correspondence. This relation can explain the biases of implied volatility regressions documented in the literature. According to this, the discrepancy between physical and risk-neutral volatilities originates from the volatility risk premium effects which can be captured by higher than second-order risk-neutral cumulants. These effects, which are priced in option markets, compensate investors for extreme negative shifts in asset markets due to, for instance, crashes or other extreme events.

Based on this theoretical relation, we suggest new regression models to investigate the information content of implied volatility or any other higher-order risk-neutral cumulant about their corresponding physical measures. These models adjust standard implied volatility regression models used in practice to account for risk premium effects. Using these models and option price data on the S&P500, FTSE100 and NASDAQ indexes, we provide strong evidence that the volatility premium can adequately capture the bias of the standard implied volatility regression to predict physical volatility. Our results indicate that this risk premium generated bias can be mainly attributed to the negative skewness of the log-return risk-neutral distributions. This can be done for relatively small values of the risk aversion coefficient. Analogous values of this coefficient are also found that can explain the risk premium biases embedded in higher than second-order risk-neutral cumulant regressions used to predict their physical counterparts.

References

- Ait-Sahalia, Y. and A.W. Lo, 2000, Nonparametric risk management and implied risk aversion, *Journal of Econometrics* 94, 9-51.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and H. Ebens, 2001, The distribution of stock return volatility, *Journal of Financial Economics*, 61, 43-76.
- Bakshi, G., N. Kapadia and D. Madan, 2003, Stock return characteristics, skew laws, and the differential pricing of individual equity options, *The Review of Financial Studies* 16, 101-143.
- Bakshi, G. and D. Madan, 2000, Spanning and derivative-security valuation, *Journal of Financial Economics* 55, 205-238.
- Bates, D., 2003, Empirical option pricing: a retrospection, *Journal of Econometrics* 116, 387-404.
- Bliss, R. and N. Panigirtzoglou, 2004, Option-implied risk aversion estimates, *Journal of Finance* 59, 407-446.
- Bollerslev, T. and H. Zhou, 2006, Volatility puzzles: a unified framework for gauging return-volatility regressions, *Journal of Econometrics* 109, 33-65.
- Canina, L. and S. Figlewski, 1993, The informational content of implied volatility, *Review of Financial Studies* 6, 659-681.
- Christensen, B.J. and N.R. Prabhala, 1998, The relation between implied and realized volatility, *Journal of Financial Economics* 50, 125-150.
- Cochrane, J. H., 2001, Asset pricing, *Princeton University Press*.
- Corrado, C.J. and T.W. Miller, 2005, The forecast quality of CBOE implied volatility indexes, *Journal of Future Markets*, 25, 339-373.
- Corrado, C. and T. Su, 1997, Implied volatility skews and stock index skewness and kurtosis implied by S&P 500 index option prices, *Journal of Derivatives* 4, 8-19.
- Davidson, R. and J.G. MacKinnon, 1993, Estimation and inference in econometrics, *Oxford University Press*.21
- Giot, P. and S. Laurent, 2007, The information content of implied volatility in light of the jump/continuous decomposition of realized volatility, *Journal of Future Markets*, 27, 337-359.
- Greene, W.H., 2004, Econometric analysis, *Prentice Hall*.
- Hansen, L. P. and K. J. Singleton, 1984, Errata: generalized instrumental variables estimation of nonlinear rational expectation models, *Econometrica* 52, 267-268.
- Jackwerth, J. C., 2000, Recovering risk aversion from option prices and realized returns, *Review of Financial Studies* 13, 433-467.
- Jiang, G. J. and Y. S. Tian, 2005, The model-free implied volatility and its information content, *Review of Financial Studies* 18, 1305-1342.
- Liu, X., M. B. Shackleton, S. J. Taylor and X. Xu, 2007, Closed-form transformations from risk-neutral to real-world distributions, *Journal of Banking and Finance* 31, 1501-1520.
- MacKinnon, J. W., H. White and R. Davidson, 1983, Tests for model specification in the presence of alternative hypotheses: some further results, *Journal of Econometrics*, 21, 53-70.
- Malevergne, Y. and D. Sornette, 2006, Extreme financial risks: from dependence to risk management, *Springer-Verlag Eds*.
- Mehra, R. and E. C. Prescott, 1985, The equity premium: a puzzle, *Journal of Monetary Economics* 15, 145-161.
- Poon, S-H. and C. Granger, 2003, Forecasting volatility in financial markets: a review, *Journal of Economic Literature* 41, 478-539.
- Rompolis, L. S. and E. Tzavalis, 2008, Recovering risk neutral densities from option prices: A new approach, *Journal of Financial and Quantitative Analysis* 43, 1037-1054.
- Rosenberg, J. and R. Engle, 2002, Empirical pricing Kernels, *Journal of Financial Economics*, 64, 341-371.
- Stuart, A. and J. K. Ord, 1994, Kendall's advanced theory of statistics (6th ed.), *New York, Halsted Press*.

Appendix

Table 1. Implied Volatility Regressions

S&P500		FTSE100		NASDAQ			
Model: $k_{\rho,2}^P = \alpha + \beta k_{\rho,2}^N + e$							
α	0.00021	(0.4E-4)		0.5E-4	(0.0003)	0.0005	(0.4E-4)
β	0.68	(0.14)		0.52	(0.09)	0.77	(0.08)
$W_{(\alpha=0,\beta=1)}$	29.52	$[pv = 0.00]$		50.63	$[pv = 0.00]$	3.13	$[pv = 0.04]$
$\overline{R}^2$	0.41			0.48		0.77	
DW	1.78			1.40		1.61	
$RMSE^{(i)}$	0.0021			0.0022		0.0033	
$MAE^{(i)}$	0.0012			0.0013		0.0021	
P_E	0.32	$[pv = 0.72]$		1.40	$[pv = 0.16]$	0.35	$[pv = 0.73]$
$RMSE^{(o)}$	0.9E-4			0.0014		0.0013	
$MAE^{(o)}$	0.8E-4			0.0007		0.0008	
Model: $\left(k_{\rho,2}^P\right)^{1/2} = \alpha + \beta \left(k_{\rho,2}^N\right)^{1/2} + e_1$							
α	0.0037	(0.0048)	0.0051	(0.0037)		0.0027	(0.0035)
β	0.77	(0.096)	0.69	(0.0740)		0.87	(0.05)
$W_{(\alpha=0,\beta=1)}$	58.58	$[pv = 0.00]$	87.67	$[pv = 0.00]$		8.91	$[pv = 0.00]$
$\overline{R}^2$	0.49		0.58			0.83	
DW	1.81		1.47			1.70	
$RMSE^{(i)}$	0.015		0.015			0.014	
$MAE^{(i)}$	0.010		0.010			0.010	
P_E	0.22	$[pv = 0.83]$	1.21			0.72	$[pv = 0.47]$
$RMSE^{(o)}$	0.0089		0.012			0.010	
$MAE^{(o)}$	0.0061		0.008			0.007	

(continued)

Table 2. Implied Volatility Regressions (continued)

	S&P500	FTSE100	NASDAQ			
	Model: $\ln(k_{\rho,2}^P)^{1/2} = \alpha + \beta \ln(k_{\rho,2}^N)^{1/2} + e_2$					
α	-0.45	(0.23)	-0.49	(0.18)	-0.20	(0.11)
β	0.92	(0.08)	0.92	(0.06)	0.95	(0.06)
$W_{(\alpha=0, \beta=1)}$	82.64	$[pv = 0.00]$	68.99	$[pv = 0.00]$	13.33	$[pv = 0.00]$
$\bar{R}^2$	0.55		0.65		0.80	
DW	1.87		1.68		1.90	
$RMSE^{(i)}$	0.27		0.20		0.18	
$MAE^{(i)}$	0.20		0.20		0.14	
$RMSE^{(o)}$	0.24		0.27		0.20	
$MAE^{(o)}$	0.18		0.19		0.14	

Notes: Table 1 presents the least squares estimates of all implied volatility regression models for the Standard & Poor's 500 (S&P500), Financial Times Stock Exchange 100 (FTSE100) and NASDAQ 100 (NASDAQ) indexes. $k_{\rho,2}^P$ is the second-order physical cumulant standing for physical volatility and $k_{\rho,2}^N$ is the second-order risk-neutral cumulant standing for risk-neutral volatility. $W_{(\alpha=0, \beta=1)}$ is the Wald test statistic of the joint hypothesis: $\alpha = 0$ and $\beta = 1$. This is distributed as χ^2 with two degrees of freedom. $\bar{R}^2$ is the adjusted coefficient of determination, $\Delta\Omega$ is the Durbin-Watson statistic for serial correlation, $RMSE^{(i)}$ and $MAE^{(i)}$ are the in-sample values of the root mean square and mean absolute prediction errors, respectively, while $RMSE^{(o)}$ and $MAE^{(o)}$ are the out-of-sample ones, covering the sample interval 2004:01-2007:12. P_E is the MacKinnon, White and Davidson test for log-linearity (see footnote 11). Standard deviations are in parentheses and p-values in brackets. The standard deviations are corrected for White heteroscedasticity and Newey-West serial correlation using four lags back.

Table 3. Implied Volatility Regressions Allowing for Risk Premium Effects

	S&P500	FTSE100	NASDAQ			
	Unrestricted Model: $k_{\rho,2}^P = \alpha + \beta k_{\rho,2}^N + \gamma k_{\rho,3}^N + e_2$					
α	-0.8E-4	(0.0005)	-0.0002	(0.0003)	0.9E-4	(0.0006)
β	0.89	(0.25)	0.92	(0.16)	0.97	(0.15)
γ	1.28	(0.89)	2.03	(0.71)	1.48	(1.10)
$W_{(\alpha=0, \beta=1)}$	1.47	$[pv = 0.47]$	5.33	$[pv = 0.07]$	0.026	$[pv = 0.97]$
$\bar{R}^2$	0.40		0.52		0.77	

DW	1.81		1.40		1.63	
$RMSE^{(i)}$	0.0020		0.0021		0.0032	
$MAE^{(i)}$	0.0011		0.0017		0.0025	
$RMSE^{(o)}$	0.0008		0.0014		0.0013	
$MAE^{(o)}$	0.0004		0.0007		0.0008	
Restricted Model ($\otimes = 1$): $k_{\rho,2}^P - k_{\rho,2}^N = \alpha + \gamma k_{\rho,3}^N + e_2^*$						
α	-0.0003	(0.0002)	-0.0004	(0.0002)	-0.3E-5	(0.0011)
γ	1.68	(0.74)	2.39	(0.34)	1.63	(0.70)
$\bar{R}^2$	0.15		0.48		0.21	
DW	1.82		1.41		1.63	
$RMSE^{(i)}$	0.0020		0.0021		0.0032	
$MAE^{(i)}$	0.0016		0.0013		0.0020	
$RMSE^{(o)}$	0.0004		0.0014		0.0012	
$MAE^{(o)}$	0.0005		0.0006		0.0008	
LR	0.35	[$pv = 0.54$]	0.44	[$pv = 0.50$]	0.052	[$pv = 0.82$]

Notes: Table 2 presents least squares estimates of both the unrestricted and restricted implied volatility regression models allowing for risk premium effects for the three indexes: Standard & Poor's 500 (S&P500), Financial Times Stock Exchange 100 (FTSE100) and NASDAQ 100 (NASDAQ). $k_{\rho,2}^P$ and $k_{\rho,3}^P$ denote the second and third-order physical cumulants, respectively, while $k_{\rho,2}^N$ and $k_{\rho,3}^N$ denote the second and third-order risk-neutral cumulants, respectively. $\Omega(\otimes=0, \otimes=1)$ is the Wald test statistic of the joint hypothesis: $\alpha = 0$ and $\beta = 1$. This is distributed as χ^2 with two degrees of freedom. $\bar{R}^2$ is the adjusted coefficient of determination, DW is the Durbin-Watson statistic for serial correlation, $RMSE^{(i)}$ and $MAE^{(i)}$ are the in-sample values of the root mean square and mean absolute prediction errors, respectively, while $RMSE^{(o)}$ and $MAE^{(o)}$ are the out-of-sample ones, covering the sample interval 2004:01-2007:12. LR is the likelihood ratio test of the restricted (under hypothesis $\beta = 1$) model against the unrestricted model, where β is estimated freely. This is distributed as χ^2 with one degree of freedom. Standard errors are in parentheses and p-values are in brackets. These are corrected for White heteroscedasticity and Newey-West serial correlation using four lags back.

Table 4. System Estimation of the Risk Aversion Coefficient $\square$

	S&P500	FTSE100	NASDAQ			
Model: $k_{\rho,n}^P - k_{\rho,n}^N = \alpha_n + \gamma_n k_{\rho,n+1}^N + e_n^*$ $v = 2, 3, 4$						
α_2	-0.0002	(0.0002)	-0.0003	(0.0002)	0.0003	(0.0002)
α_3	0.0001	(0.2E-0)	0.0002	(0.2E-0)	0.0003	(0.0002)
α_4	-0.2E-4	(0.2E-0)	-0.4E-4	(0.2E-0)	-0.8E-4	(0.2E-0)
γ_2	1.78	(0.71)	2.57	(0.28)	2.10	(0.77)
γ_3	2.83	(0.15)	2.58	(0.19)	2.68	(0.29)

γ_4	2.86	(0.16)	3.03	(0.20)	3.15	0.45
$W_{(\gamma_2=\gamma_3)}$	1.75	[$pv = 0.18$]	0.0006	[$pv = 0.98$]	0.29	[$pv = 0.58$]
$W_{(\gamma_2=\gamma_4)}$	1.89	[$pv = 0.17$]	1.14	[$pv = 0.28$]	0.67	[$pv = 0.41$]
$W_{(\gamma_2=\gamma_3=\gamma_4)}$	2.33	[$pv = 0.31$]	26.99	[$pv = 0.00$]	0.29	[$pv = 0.90$]

Notes: Table 3 presents estimates of the equation system containing the second, third and fourth-order cumulant restricted regressions. The system is estimated based on the SUR method for the following three indexes: Standard & Poor's 500 (S&P500), Financial Times Stock Exchange 100 (FTSE100) and NASDAQ 100 (NASDAQ). $k_{\rho,n}^P$ denotes the n -order physical cumulant and $k_{\rho,n}^N$ denotes the n th-order risk-neutral cumulant. $W_{(\gamma_2=\gamma_3)}$ and $W_{(\gamma_2=\gamma_4)}$ are the Wald test statistics of the joint hypotheses $\gamma_2=\gamma_3$ and $\gamma_2=\gamma_4$, respectively. These are distributed as χ^2 with two degrees of freedom. $W_{(\gamma_2=\gamma_3=\gamma_4)}$ is the Wald test statistic for the joint hypothesis $\gamma_2=\gamma_3=\gamma_4$. This is distributed as χ^2 with two degrees of freedom. Standard errors are in parentheses and p-values in brackets. These are corrected for White heteroscedasticity and Newey-West serial correlation using four lags back.