Master Thesis
Quantitative Finance
Erasmus School of Economics
Department of Econometrics

Return Predictability from Realized and Option-Implied Moments

Evidence from the S&P 500 index

Author:
Jim Kaandorp (336473)

Supervisor:
Prof. Dr. Dick. J. van Dijk

Co-reader:
Dr. Bart F. Diris

September 30, 2011

Abstract

In this paper we test which computational method, of computing statistical moments, yields time series that contain most information on future equity returns. The methods we consider are the realized, autoregressive forecasted realized (AR), and option-implied method, the latter both model-free and derived from a Generalized Beta Distribution of the Second Kind ($GB2^{Q}$). Using a static time series regression, we find a significant positive relation between the volatility of the AR^{315d} (estimated on 315 trading-days), model-free, and $GB2^{Q}$ method and subsequent excess returns. We do not find any relation using the skewness, and kurtosis. The rolling window regressions show that the coefficients vary substantially over time. Furthermore, these results indicate that the higher moments do not improve the accuracy of forecasting the excess return, and that we lack the statistical support to prefer one method over another in terms of forecasting accuracy. In measuring the performance of trading strategies, created using an utility framework, we find that the model-free and AR^{315d} method perform best, with Sortino ratios of 0.30, compared to 0.24 of the S&P 500 index. Second best is the $GB2^{Q}$ method, followed by the realized and AR^{15m} method (estimated on 15 months). Measured in alpha, the AR^{315d} method performs better than the model-free method. The performance of the daily AR method is enhanced when increasing the estimation window when estimating the AR coefficients. Imposing an upper bound constraint on the optimal weight, slightly improves the performance of all methods. Furthermore, we find that the contribution of the third and fourth moment are economically not significant, that none of the methods perform consistently over time, and that the performance is insensitive to changes in the relative risk aversion parameter.

Keywords: return-predictability, realized volatility, option-implied volatility, higher moments, utility framework.

Table of Contents

1 Introduction

Predictability of equity returns is a subject that is of key interest to investors and researchers. In the past, some variables were found to have a relation with future returns. Most common examples are the level and spread of interest rates, dividend yield, and price-earnings ratio. More recently, also statistical moments of the return distribution entered the equation. Some papers indicate that those moments are relevant in explaining future returns. For example, Amaya et al. (2011) find explanatory power of realized skewness and kurtosis. Just as with the second moment, the higher moments describe the shape of the return distribution and can therefore be viewed as 'risk-factors.' Hence, these higher moments should be related to the required rate of return on an investment. Also other methods to compute such moments are used. For example, Conrad et al. (2009) use option-implied moments in their analysis.

Based on literature, statistical moments seem to contain information on future returns. However, given that there are numerous ways to derive these moments it is not clear which method to compute these moments performs best, both from a statistical and economic point of view. Furthermore, there are still methods that have not been investigated in such a forecasting setting before, like the method using the Generalized Beta Distribution of the Second Kind (GB2). This distribution is often used to describe the return distribution as it is rather flexible in its shape, yet it has only three parameters to estimate.

Most papers analyze the cross-section of returns, whereas a time series setting is more relevant for the capital allocation decision, stocks versus cash. We focus on the time series relation, since the capital allocation decision generally has a much larger influence on the final returns than the security selection decision.

Along these lines, we formulate the following research question: Which computational method yields statistical moments that contain most information about future excess returns? In answering this question we formulate the following sub-questions: i) Is there a relation between the statistical moments of different computational methods and future excess returns? ii) What is the forecasting power of the statistical moments in predicting future excess returns? iii) What is the contribution of the individual moments to this forecasting power? iv) What is the economic performance of trading strategies based on the statistical moments? v) What is the contribution of the individual moments to this performance?

We consider the S&P 500 index for this analysis as proxy for the aggregate market portfolio, from January 1996 to December 2010. We limit ourselves to the second, third, and fourth standardized and non-standardized moments.

In total we use five different methods to compute the statistical moments. These are the realized method, forecasted realized method based on a daily and monthly autoregressive (AR) structure, model-free option-implied risk-neutral method, and parametric option-implied risk-neutral GB2 method ($GB2^{Q}$). We expect most from the 'forward looking' methods, as we expect financial markets to be forward looking as well. The AR methods are forward looking, as the statistical moments are forecasted. According to Battalio and Schultz (2006), also option-implied moments are a true *ex ante* measure of expectations, when option prices reflect the same information as stock prices. Bates (1991), Jackwerth and Rubinstein (1996), and Rubinstein (1985, 1994) postulate that option markets appear to be efficient in incorporating the information of market participants. Hence, also the option-implied method is compelling to include in the analysis.

We test the information content of the moments in a linear regression setting as well as an utility framework. Using static regressions, we regress the future excess returns on the moment variables and test for significant coefficients. Using rolling window regressions, we analyze the accuracy in predicting future excess returns, out-of-sample. To avoid spurious results, other explanatory variables are included in the regressions. Furthermore, we use a rolling window in the out-of-sample regressions to account for possible structural breaks, and analyze the parameter variability over time.

Concerning the utility framework, we create trading strategies which are analyzed in terms of their economic performance. A trading strategy consist of two assets, the S&P 500, and the 'risk-free' one-month T-bill rate. A utility framework is used to determine the optimal portfolio weight in the S&P 500, each month, by using the information contained in the statistical moments. In this way, we test if the moment series contain information on future excess returns, via the predictability of the moment variables themselves. Note that their should still be a relation between the moment series and future excess returns to gain outperformance. Otherwise the adjustments in the portfolio weight will not be rewarded, and hence the performance does not improve. This relation does not have to be linear, however.

With the static regressions we find a significant relation between the volatility of the AR^{315d}, model-free, and $GB2^{Q}$ method and subsequent excess returns. We do not find any relation using the skewness and kurtosis. The rolling window regressions show that the coefficients vary substantially over time. Furthermore, we find some support for the volatility variable in improving the accuracy of the forecasts. When comparing the accuracy of the different models, we lack the statistical support to prefer one method over another.

Using the utility framework, we find that the $\text{AR}^{315\text{d}}$ and model-free method perform best with Sortino ratio's of 0.30, compared to a ratio of 0.24 for the S&P 500. Second best is the GB2^{Q} method, followed by the realized, and $\text{AR}^{15\text{m}}$ method, with ratios equal to 0.26, 0.19, and 0.18 respectively. Measured in alpha, the $\text{AR}^{315\text{d}}$ method performs better than the model-free method. The alphas are respectively 1.02, 0.65, 0.43, 0.29, and 0.28 percent annually for the $\text{AR}^{315\text{d}}$, model-free, GB2^{Q} method, realized and $\text{AR}^{15\text{m}}$ method. Effectively all outperformance can be attributed to the second moment; the contribution of the third and fourth moment are economically not significant.

Non of the strategies perform consistently over time. Measured halfway 2006, the M^2 measure of all strategies was negative. Hence, the positive overall M^2 measure is due to the last three and a half years of our sample period. Measured over different values of the relative risk aversion parameter, the performance is stable.

The performance of all strategies improve moderately when imposing an upper bound restriction on the optimal weight. When the restriction gets too stringent, the performance deteriorates. Looking at the Sortino ratio, we find maximum improvements of 0.01, 0.04, 0.02, 0.02, and 0.04 for the realized, $\text{AR}^{315\text{d}}$, $\text{AR}^{15\text{m}}$, model-free, and GB2^{Q} method respectively. The alphas improve respectively by a maximum of 0.04, 0.11, 0.02, 0.00, and 0.10 percent annually.

When testing for different estimation lengths of the AR models, we find that for the daily model the performance gradually improves when the estimation window increases. For the monthly model, the performance slightly improves when the estimation window gets shorter. Overall, the daily model performs better than the monthly model. The best performing AR model yields a Sortino ratio and alpha of respectively 0.32 and 1.13 percent, thereby outperforming the model-free method on both measures.

With this paper we contribute to the existing literature in multiple ways. As a first contribution, we make a comparison between five methods to compute statistical moments. Other papers like Amaya et al. (2011), Ammann et al. (2009), and Conrad et al. (2009), only consider one or two methods.

The second contribution is that we use two methods that, according to our knowledge, have not been used before in such a forecasting setting. These are the parametric GB2 distribution and the AR structure to forecast realized moments. The GB2 distribution has been used as a way to extract the probability density function from option prices, however, these are not used in forecasting returns. (see for example: Bookstaber and Mc Donald (1987,1991)). Jondeau and Rockinger (2008) also forecast moments, however, implicitly by deriving them from a forecasted return distribution using a dynamic conditional correlations model.

The third contribution is that we use a much wider range of methods to evaluate the statistical and economical performance. Next to static Ordinary Least Squares (OLS) regressions, as used in most papers in the form of Fama and French (1995) regressions, we also perform out-of-sample rolling regressions and use an utility framework to create trading strategies.

As a final contribution, we focus on the time series of returns of one equity index, and hence on the capital allocation decision. Other papers use multiple individual stocks and often focus on the cross-section of returns. (see for example: Conrad et al. (2009), Amaya et al. (2011), Ammann et al. (2009) Xing and Zhou (2010)).

The findings in this paper are of interest to both practitioners and academics. Amongst the practitioners we should think of active asset managers, like hedge-fund and mutual-fund managers. For academics, this research is interesting for the contributions as stated above.

The remainder of this paper proceeds as follows: Section 2 provides a description of the literature. Section 3 discusses the data. Section 4 describes the methodology of i) computing the statistical moments, and ii) the testing procedures we use. Section 5 discusses the empirical results. Finally, Section 6 concludes.

2 Literature Review

This section discusses the literature related to testing the relation between statistical moments and future equity returns. We discuss the results of papers that use realized and option-implied moments. Moreover, relevant literature is discussed concerning the utility framework.

Relating to the realized moments, Amaya et al. (2011) use Fama and French (1995) time series regressions to decile portfolios, as well as cross-sectional Fama and MacBeth (1973) regressions, to test for a relation between weekly realized volatility, skewness, and kurtosis and subsequent returns. Model-free methods based on Anderen et al. (2001) are used to compute the moments. They find no clear relation between realized volatility and future stock returns. However, they do find a strong negative relation using realized skewness and a strong positive relation using realized kurtosis. When they use double-sort portfolios the relation becomes weaker. The results are robust over different sub-samples and various control variables.

Their findings concerning the skewness and kurtosis make sense when assuming there is a risk-return trade-off in place and that markets are efficient in this respect. When risk is defined as the probability times the magnitude of a loss, a higher probability of extreme returns (high kurtosis) is more risky. Also, a higher probability of (high) positive returns relative to (high) negative returns (positive skewness) is less risky. When investors are risk-averse, a higher risk demands a higher risk premium. For the skewness, this idea is formalized by Harvey and Siddique (2000), who incorporate co-skewness into a traditional mean-variance asset pricing model. They show that including a skewness factor (based on historical returns) in the Fama-French model applied to decile portfolios, vastly reduced the F-statistic on the significance of the intercept (alpha). They also show that momentum, as analyzed by Carhart (1997), is to a large extent related to systematic skewness. Hence, they postulate that co-skewness is an important determinant in explaining the variation of equity returns of individual US stocks. Consistent with our intuition, they find that a positive skewness comes along with a higher return.

Conrad et al. (2009) study the relation between option-implied moments and future stock returns. They also test for a significant intercept (alpha) by applying Fama-French regressions to decile portfolios. This is done for 1, 3, 6, and 12 months time-to-expiration options, on all stocks as documented by the OptionMetrics database. A model-free method to extract moments from option-prices is used, based on quadratic, cubic, and quartic returns, as first proposed by Bakshi et al. (2003). They find a negative relation between the volatility and next month's returns. Also, for the skewness they find a negative relation, with the exception of one month time-to-expiration options. Furthermore, they find a positive relation between the kurtosis and future returns. The relations are both economically and statistically significant (at a 10 percent level).

Ammann et al. (2009) use the implied volatilities of individual US stocks in their analysis, as provided by the OptionMetrics database. They run panel-data regressions, in which the future monthly excess return is regressed on the implied volatility. In this way they test for a significant variable coefficient instead of intercept. Their main finding is a highly significant, positive relation between implied volatilities and subsequent returns. These findings seem to contradict the findings of Conrad et al. (2009). Given that they both use data from 1996 to 2005, the differences in results are likely due to the different methodologies used. Conrad et al. (2009) test for an intercept after all.

Jondeau and Rockinger (2008) research the effect of what they call 'distributional timing' in portfolio allocation. They do so by comparing the performance of trading strategies based on forecasted historical volatility, skewness, and kurtosis to their non-forecasted (lagged) counterparts, using an utility framework. They use a dynamic conditional correlations model to forecast the distribution of returns, from which the moments are derived. The utility framework is used to allocate weights to either US, Japanese, and UK stocks as well as a risk-free asset, all on a weekly basis. Their main finding is that distributional timing adds economic value, with an outperformance of 1.40 percent per year. Volatility alone accounts for 0.55 percent per year.

3 Data

The option prices are retrieved from the OptionMetrics database. Since the methods used to compute the option-implied moments are based on European style options, and before 1996 only American options were traded, our data set 'only' spans the period from January 1996 to March 2010. The time-to-expiration of the options equals 30 days. This is in accordance with the time-to-expiration of the Chicago Board Options Exchange Volatility Index (VIX).

We make use of closing prices, computed as the midpoint average of the closing bid and ask prices. To ensure we use accurate prices, we eliminate options with a volume lower than 100 contracts per day and with a closing price lower than 0.5 USD. We also check for no-arbitrage violations. Strictly checking the put-call parity would result in too many eliminations, as not all options are traded exactly at the same time. Still, we check for the following upper and lower bounds that the bid and ask prices should satisfy:

$$c_t^b(X,\tau) \leq P_t\, e^{-q\tau} \ , \tag{1}$$

$$c_t^a(X,\tau) \geq \ \max\{P_t\, e^{-q\tau} - X\, e^{-r_f\tau}, 0\} \ , \tag{2}$$

$$p_t^b(X,\tau) \leq X\, e^{-r_f\tau} \ , \tag{3}$$

$$p_t^a(X,\tau) \geq \ \max\{X\, e^{-r_f\tau} - P_t\, e^{-q\tau}, 0\} \ , \tag{4}$$

where $c_t(X,\tau)$ and $p_t(X,\tau)$ are either the bid (b) or ask (a) prices of respectively a European call and put option, with strike-price X and time-to-expiration τ. P_t and q are respectively the price and dividend yield of the S&P 500, and r_f the the 'risk-free' one-month T-bill rate.

We use Tickdata[1] as source for the high-frequency data of the S&P 500 index, from which the realized moments are computed. This data is available from January 1990 to March 2010. It is aggregated to five minute returns. For the S&P 500, this is considered the optimal sampling frequency according to studies like Andersen et al. (1999) and Hansen and Lunde (2006). Increasing the sampling frequency could introduce noise, as a result of the market microstructure like bid-ask spreads.

The returns are computed from a total return index. This index is retrieved from the Datastream database. Also the series to construct the control variables are obtained from this database. These are the price-earnings ratio and dividend-yield of the S&P 500, as well as

[1] See www.tickdata.com

the one and three-month T-bill rate, the 10 year T-bond yield, and Moody's AAA and BAA corporate bond credit spread.

4 Methodology

This section is divided in two parts. The first part discusses the methods used to obtain the statistical moments. The second part discusses the methods used for analyzing the predictive content, both statistically and economically. In the discussion that follows, the S&P 500 represents the 'underlying,' and the one-month T-bill rate the 'risk-free rate.'

4.1 Computational Methods

4.1.1 Realized Moments

The definitions of the non-standardized moments are derived from the general definition of power variation as stated in Barndorff and Shephard (2003). Using this definition results in model-free measures that converge to the integrated moment limits when the sampling frequency increases. With the intra-day return, $r_{t,i}$, defined as

$$r_{t,i} = \ln(P_{t-1+\frac{i}{N}}) - \ln(P_{t-1+\frac{i-1}{N}}) \ , \tag{5}$$

with P_l the price of the S&P 500, at time l, and $N + 1$ the total number of observations per trading day, this leads to

$$m_{2,t} = \sum_{i=1}^{N} r_{t,i}^2 \ , \tag{6}$$

$$m_{3,t} = \sqrt{N} \sum_{i=1}^{N} r_{t,i}^3 \ , \tag{7}$$

$$m_{4,t} = N \sum_{i=1}^{N} r_{t,i}^4 \ , \tag{8}$$

for the second, third, and fourth non-standardized (sample) moments. The standardized moments follow directly from their definitions. Hence, the realized volatility, skewness, and kurtosis are respectively defined as

$$v_t = m_{2,t} = \sum_{i=1}^{N} r_{t,i}^2 \,, \tag{9}$$

$$s_t = \frac{m_{3,t}}{\sqrt{(m_{2,t})^3}} = \frac{\sqrt{N}\sum_{i=1}^{N} r_{t,i}^3}{(\sum_{i=1}^{N} r_{t,i}^2)^{3/2}} \,, \tag{10}$$

$$k_t = \frac{m_{4,t}}{(m_{2,t})^2} = \frac{N\sum_{i=1}^{N} r_{t,i}^4}{(\sum_{i=1}^{N} r_{t,i}^2)^2} \,, \tag{11}$$

which is consistent with Amaya et al. (2011).

Since we use a monthly horizon, we construct monthly realized moments by averaging the daily moments over 21 trading days. Hence, the moments as used in the utility framework, being the realized monthly second (RMM_2), third (RMM_3), and fourth (RMM_4) moment, are computed as[2]

$$RMM_{2,t} = \frac{1}{21}\sum_{i=0}^{20} m_{2,t}, \tag{12}$$

$$RMM_{3,t} = \frac{1}{21}\sum_{i=0}^{20} m_{3,t} \,, \tag{13}$$

$$RMM_{4,t} = \frac{1}{21}\sum_{i=0}^{20} m_{4,t} \,, \tag{14}$$

with, t, in trading-days. The moments as used in the regressions, being the the realized monthly volatility (RMV), skewness (RMS), and kurtosis (RMK), are computed as

$$RMV_t = \left(\frac{252}{21}\sum_{i=0}^{20} v_{t-i}\right)^{1/2} \,, \tag{15}$$

$$RMS_t = \frac{1}{21}\sum_{i=0}^{20} s_{t-i} \,, \tag{16}$$

$$RMK_t = \frac{1}{21}\sum_{i=0}^{20} k_{t-i} \,, \tag{17}$$

[2]Alternatively, we leave out the first half hour of each trading day, when computing the realized (standardized) moments. This could yield a better measure, as the firsts half hour is generally characterized by relatively large deviations in returns (especially the first five minute interval). The results when using this alternative measure, are provided by Section 7.2.2 in the Appendix.

with, t, in trading-days. Note that the volatility variable is annualized.

4.1.2 Autoregressive Forecasts

We forecast the realized moments one month ahead, consistent with the time-to-maturity of the options. Two different AR models are used for this purpose. One model fitted on daily observations and one model fitted on monthly observations, of the monthly measures as defined in equation 12 to 17. The monthly model is used in addition to the daily model, as the information contained in the daily observations could not be relevant in forecasting one month ahead. If so, these (intermediate) daily observations could introduce noise and thereby reducing the forecasting power.

To ensure positive forecasts for both the second and fourth moment, we forecast the natural logarithm of these variables. By assuming the second and fourth moment to be lognormally distributed, we can derive the non-transformed moments from the forecasted log-moments using the following transformation:

$$x_t = e^{\ln(x_t)+\frac{1}{2}V(\ln(x_t))} , \tag{18}$$

with x_t being either the second or fourth (standardized) moment. The AR models are fitted directly to the time series of the third moments. Hence, these are not transformed.

The AR models are defined as

$$y_t = \mu + \sum_{i=1}^{K} \phi_i y_{t-i} + \varepsilon_t, \tag{19}$$

with y_t being equal to $\ln(x_t)$, in case of the second and fourth moment, and equal to the (standardized) moment itself in case of the third (standardized) moment. The variable ε_t is an i.i.d. innovation process with E[ε_t] = 0 and K the number of AR lags. The subscript, t, is either in trading-days or in months. Note that V($\ln(x_t)$) in the transformation is equal to the variance of the error terms of the AR model, V(ϵ_t).

We set K and the estimation window in such a way that the Mean Squared Prediction Error (MSPE) of predicting the (transformed) moments is minimized over the period from July 1992 to January 1996 (excluding estimation period). Table 12 and 13 in the Appendix provide the MSPE's for different values of K and lengths of the estimation window, for the

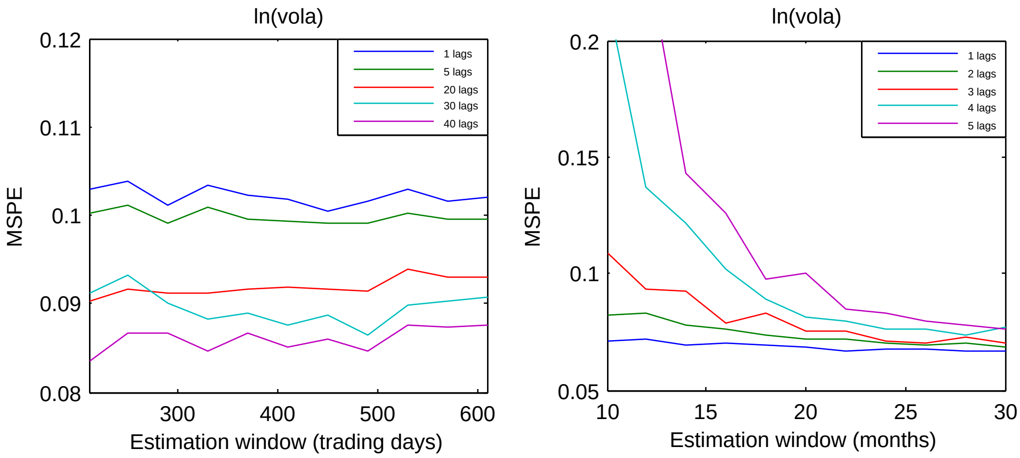

Figure 1: MSPE's of predicting the log-volatility
This figure displays the MSPE's of the daily (left) and monthly (right) AR model, fitted on the log-volatility series, for different values of K and lengths of the estimation window.

Table 1: AR lag specification
This table shows the number of lags, K, as used in the daily and monthly AR model, for the different (log-transformed) moment series.

	Daily Model	Monthly Model
$\ln(m_2)$	40	1
m_3	1	1
$\ln(m_4)$	40	1
ln(vola)	40	1
skewness	1	1
ln(kurt)	40	2

daily and monthly models respectively.

As an example, we graphically show these results for the log-volatility in Figure 1. Clearly, for the daily model a lag of 40 days yields the lowest MSPE, whereas for the monthly model one lag is preferred. Table 1 provides an overview of the lags selected for all models.

Relating to the length of the estimation window, we find that the curve pertaining to a lag of 40 days steepens somewhat, whereas it declines moderately for the one month lag, when the estimation length increases. However, these slopes are marginal, suggesting that there is a relatively wide range of lengths that is optimal. We find similar patterns with the remaining moments. For consistency, we want the lengths of all models to be the same and hence we compromise on a estimation window of 15 months (or 315 trading-days) for our base case scenario. As a robustness check, we also compute the results for window sizes varying from 10 to 25 months.

4.1.3 Option-Implied Moments

We use both a model-free and parametric method to compute the statistical moments. Since both approached are completely different, we can asses whether the performance of the option-implied moments depend on how the moments are derived. The model-free method allows for more flexibility in especially the higher moments. This flexibility is not better per se, as it could also introduce noise in the time series. Next, both methods are discussed in detail.

Model-Free Method

The model-free approach relies on a finding by Bakshi et. al. (2003), who show that the price of a security that pays the quadratic, cubic, and quantic return on a base security can be expressed as functions of the prices of call and put options. According to their work, the prices of these payoffs can be expressed as

$$V_t(\tau) = \int_{P_t}^{\infty} \frac{2(1-\ln(X/P_t))}{X^2} c_t(X,\tau)\mathrm{d}X + \int_{0}^{P_t} \frac{2(1+\ln(X/P_t))}{X^2} p_t(X,\tau)\mathrm{d}X \ , \tag{20}$$

$$W_t(\tau) = \int_{P_t}^{\infty} \frac{6\ln(X/P_t) - 3(\ln(X/P_t))^2}{X^2} c_t(X,\tau)\mathrm{d}X + \int_{0}^{P_t} \frac{6\ln(X/P_t) + 3(\ln(X/P_t))^2}{X^2} p_t(X,\tau)\mathrm{d}X \ , \tag{21}$$

$$X_t(\tau) = \int_{P_t}^{\infty} \frac{12(\ln(X/P_t))^2 - 4(\ln(X/P_t))^3}{X^2} c_t(X,\tau)\mathrm{d}X + \int_{0}^{P_t} \frac{12(\ln(X/P_t))^2 + 4(\ln(X/P_t))^3}{X^2} p_t(X,\tau)\mathrm{d}X \ , \tag{22}$$

with $V_t(\tau)$, $W_t(\tau)$, and $W_t(\tau)$ the prices of a security paying the quadratic, cubic, and quantic payoff respectively, at time t. With $c_t(X,\tau)$ and $p_t(X,\tau)$ representing the prices of a European call and put option respectively, with strike-price X and time-to-expiration τ, and P_t the price of the underlying.

We approximate the security prices in equation 20 till 22 using trapezoidal integration. Hence, the approximated security prices equal a weighted sum of out-of-the-money call and put options. We use these security prices, in turn, to derive the risk-neutral volatility, V_t^Q, skewness, S_t^Q, and kurtosis, K_t^Q, of the S&P 500, by working from their definitions. Hence,

$$V_t^Q(\tau) = E_Q[(r_{sp} - \mu)^2] = e^{r_f \tau} V_t(\tau) - \mu_t(\tau)^2 \ , \tag{23}$$

$$S_t^Q(\tau) = \frac{E_Q[(r_{sp} - \mu)^3]}{E_Q[(r_{sp} - \mu)^2]^{3/2}} = \frac{e^{r_f \tau} W_t(\tau) - 3\mu_t(\tau) e^{r_f \tau} V_t(\tau) + 2\mu_t(\tau)^3}{[e^{r_f \tau} V_t(\tau) - \mu_t(\tau)^2]^{3/2}} \ , \tag{24}$$

$$K_t^Q(\tau) = \frac{E_Q[(r_{sp} - \mu)^4]}{E_Q[(r_{sp} - \mu)^2]^2} = \frac{e^{r_f \tau} X_t(\tau) - 4\mu_t(\tau) W_t(\tau) + 6 e^{r_f \tau} \mu_t(\tau)^2 V_t(\tau) - \mu_t(\tau)^4}{[e^{r_f \tau} V_t(\tau) - \mu_t(\tau)^2]^2} \ , \tag{25}$$

with

$$\mu_t(\tau) = e^{r_f \tau} - 1 - e^{r_f \tau} V_t(\tau)/2 - e^{r_f \tau} W_t(\tau)/6 - e^{r_f \tau} X_t(\tau)/24 \ . \tag{26}$$

Parametric Method

Following Bookstaber and McDonald (1987 and 1991), we fit a GB2 distribution on the option data to obtain the entire risk-neutral density. Hence, consider the formula to price an European call option[3],

$$c_t(X,\tau) = e^{-r_f \tau} \int_X^\infty (P_T - X) f(P_T \mid \theta) \mathrm{d}P_T \ , \tag{27}$$

with $f(P_T \mid \theta)$ the risk-neutral probability density function of the price of the underlying, P_T, at time of expiration, T, with parameters θ, and r_f the risk-free rate. Furthermore, τ is the

[3]See for example Baxter and Rennie (2008) for derivation.

time-to-expiration, which is equal to $T - t$. In our case $f(P_T \mid \theta)$ equals the GB2 probability density function, defined as

$$f_{GB2}(x \mid \alpha, \beta, \rho, \nu) = \frac{\alpha}{\beta^{\alpha\rho} B(\rho, \nu)} \frac{x^{\alpha\rho - 1}}{[1 + (x/\beta)\alpha]^{\rho+\nu}} , \tag{28}$$

with $B(.)$ being defined in terms of the gamma function, as $B(\rho, \nu) = \frac{\Gamma(\rho)\Gamma(\nu)}{\Gamma(\rho+\nu)}$[4]. This distribution has four parameters, however, β is imposed by the others in the following way[5]:

$$\beta = \frac{F_t B(\rho, \nu)}{\beta B(\rho + 1/\alpha, \nu - 1/\alpha)} , \tag{29}$$

with F_t the futures price of the underlying, defined as $F_t = P_t e^{(r_f - q)\tau}$, with q the dividend yield. This leaves only three parameters to estimate.

The cumulative GB2 distribution can be evaluated by using the cumulative beta distribution, defined as

$$F_\beta(u \mid \rho, \nu) = \frac{1}{B(\rho, \nu)} \int_0^u t^{\rho-1}(1-t)^{\nu-1} \mathrm{d}t . \tag{30}$$

A change of variable procedure shows that these distributions are related in the following way:

$$F_{GB2}(x \mid \alpha, \beta, \rho, \nu) = F_\beta(u(x, \alpha, \beta) \mid \rho, \nu) . \tag{31}$$

with the function u defined as $u(x, \alpha, \beta) = \frac{(x/\beta)^\alpha}{1+(x/\beta)^\alpha}$. By substituting the right-hand-side of equation 31 into equation 27, and solving the integral, we obtain the following closed-form solution for the call price:

$$\begin{aligned} c_t(X, \tau) = F_t e^{-r_f \tau} [\, 1 - F_\beta(u(X, \alpha, \beta) \mid \rho + \alpha^{-1}, \nu - \alpha^{-1} \,] \\ - X e^{-r_f \tau} [\, 1 - F_\beta(u(X, \alpha, \beta) \mid \rho, \nu) \,] . \end{aligned} \tag{32}$$

The risk-neutral price distribution can be extracted from option prices by estimating the free parameters of the beta distribution. The objective in this optimization is to minimize the

[4]Note that the gamma function is defined as

$$\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} \mathrm{d}t .$$

[5]See Taylor (2007) for detailed explanation.

sum of squared differences between the model and market call prices. Hence,

$$\min_{\{\theta\}} \text{SSE} = \sum_{i=1}^{N} (c_{market}(X_i, \tau) - c_{model}(X_i, \tau \mid \theta_i))^2 \ . \tag{33}$$

Finally, this risk neutral price distribution is transformed to the risk-neutral return distribution. The statistical moments are derived from the latter.

Put options are not considered in this optimization. The reason is that there put options for values of X that are not covered by a call options, and which do satisfy our filter criteria are rare. These put options are all far out-of-the money, hence illiquid and priced below our threshold of 0.5 USD.

4.2 Evaluation Methods

4.2.1 Ordinary Least Squares Regression

We regress the return in excess of the risk-free rate on several explanatory variables, being the log-volatility, skewness, and log-kurtosis, as well as six control variables. In all regressions, Newey and West (1986) standard errors are used to account for heteroskedasticity and serial correlation. The regression equation is specified as

$$r_{e,sp,t} = \alpha + \beta_1' \text{moments}_{t-1} + \beta_2' \text{controls}_{t-1} + \varepsilon_t \ , \tag{34}$$

$$\text{moments}_t = (\ln(VOLA_t), SKEW_t, \ln(KURT_t))' \ , \tag{35}$$

$$\text{controls}_t = (\ln PE_t, DY_t, SD1M_t, 3M_t, DS_t, TS_t)' \ , \tag{36}$$

where $r_{e,sp,t}$ is the excess return on the S&P 500, at month t, and ε_t an i.i.d. innovation process with E$[\varepsilon_t]$ = 0. The input to the vector, moments$_t$, are the log-volatility ($\ln(VOLA_t)$), skewness, ($SKEW_t$), and log-kurtosis ($\ln(KURT_t)$), and depend on which computational method is used. Furthermore, $\ln(PE_t)$ represents the log-price-earnings ratio and DY_t the dividend yield, of the S&P 500. $SD1M$ is the stochastically de-trended one-month T-bill (yield in deviation from its 12 month moving average), DS_t the default spread (yield on Moody's BAA rated corporate bonds minus the yield on the AAA corporate bonds), and TS_t the term spread (yield on the 10 year T-bond minus the yield on the 3 month T-bill). The coefficients contained in vector β_1 are the ones we analyze. These control variables are consistent with those of Bollerslev et al. (2009) when they test for the significance the volatility-risk-premium, a 'volatility-like' variable. For an extensive discussion on these and other predictive variables, we refer to Welch and Goyal (2008).

Complementary to the static in-sample regressions, we use out-of-sample rolling window regressions to test for the accuracy of the excess return forecasts. The MSPE is used as loss function in this analysis. Applying this to the unrestricted models provides us with a comparison between the different methods, for all three moments. We also try to determine the contribution of the individual moments to the accuracy of the forecasts, by running both restricted and unrestricted regressions for each method. Three restricted versions are used, one with only the kurtosis coefficient restricted, one with the skewness and kurtosis coefficient restricted and one with all moment coefficients restricted.

Since the fourth and second (standardized) moment are dependent by definition, we expect a non-negative correlation between these variables[6]. Therefore, we also perform regressions when only including a single moment variable in the regression (next to the control variables), as opposed to including all three. Additionally, the F-statistic is reported on the joint significance of all three variables.

We use a rolling window to account for any structural breaks and compute the results for multiple window lengths, being 20, 40, and 60 months. A smaller window-size than 20 results in too large estimation errors, given that we have 9 variables to estimate. A window length larger than 60 results in less flexibility in dealing with possible structural breaks.

4.2.2 Utility Framework

In developing the utility framework, we follow Jondeau and Rockinger (2008). They use a Taylor approximation of the power utility function, a theory based on a earlier working paper of Harvey et al. (2010) and Guidolin and Timmermann (2008). So, consider that we want to maximize our expected utility, $\mathrm{E}_t[U(W_{t+1})]$, with W_{t+1} being our wealth at month $t+1$. This wealth depends on our current wealth, W_t, and the (simple) return on the investment portfolio, $r_{p,t+1}$, in the following way: $W_{t+1} = W_t(1 + r_{p,t+1})$. For convenience we set W_t equal to one.

[6]This correlation is often found negative, as kurtosis is typically low when volatility is high and vise versa.

Given that the portfolio consists of two assets, the portfolio return can be expressed as $r_{p,t+1} = \omega_t \, r_{sp,t+1} + (1-\omega) \, r_{f,t}$, with $r_{sp,t+1}$ the return of the S&P 500, $r_{f,t}$ the 'risk-free' one-month T-Bill rate, and ω_t the weight in the S&P 500.

The portfolio weight ω is obtained by maximizing the expected utility,

$$\max_{\{\omega_t\}} \mathrm{E}_t[U(W_{t+1}(\omega_t))] = \mathrm{E}_t[U(\omega_t \, r_{sp,t+1} + (1-\omega) \, r_{f,t})] \; , \tag{37}$$

$$s.t. \; \omega \; \leq c \; , \tag{38}$$

with E_t the expectation, conditional on all information up to time t, and c the upper bound constraint on ω, which equals ∞ in our base case scenario. When a Taylor approximation to the utility function is used, we get

$$U(W^{t+1}) = \sum_{k=0}^{\infty} \frac{U^{(k)}(W_t)}{k!} (W_{t+1} - W_t)^k \; , \tag{39}$$

with $W_{t+1} - W_t = r_{sp,t+1}$, the return of the portfolio, and $U^{(k)}$ the k^{th} derivative of the utility function. By assuming the utility function and its derivatives to be constant, we can run the expectation operator trough the formula in the following way:

$$\mathrm{E}_t[U(W_{t+1})] = \mathrm{E}_t \left[\sum_{k=0}^{\infty} \frac{U^{(k)}(W_t)}{k!} (r_{p,t+1})^k \right] = \sum_{k=0}^{\infty} \frac{U^{(k)}(W_t)}{k!} \mathrm{E}_t \left[(r_{p,t+1})^k \right] \; . \tag{40}$$

By limiting ourselves to the first four moments, and by noting that $\mathrm{E}_t \left[(r_{p,t+1})^k \right] = m_{p,t+1}^{(k)}$, the k^{th} non-standardized moment of the portfolio, we can approximate the expected utility as

$$\begin{aligned} \mathrm{E}_t[U(W_{t+1})] \approx U(W_t) + U^{(1)}(W_t) m_{p,t+1}^{(1)} + \frac{U^{(2)}(W_t)}{2} m_{p,t+1}^{(2)} \\ + \frac{U^{(3)}(W_t)}{3!} m_{p,t+1}^{(3)} + \frac{U^{(4)}(W_t)}{4!} m_{p,t+1}^{(4)} \; . \end{aligned} \tag{41}$$

When using the power utility function $U(W_{t+1}) = \frac{W_{t+1}^{1-\gamma}}{1-\gamma}$, with $\gamma > 1$ the relative risk aversion parameter, the expression for the expected utility finally becomes

$$\mathrm{E}_t[U(W_{t+1})] \approx \frac{1}{1-\gamma} + m^{(1)}_{p,t+1} - \frac{\gamma}{2} m^{(2)}_{p,t+1} + \frac{\gamma(\gamma+1)}{3!} m^{(3)}_{p,t+1} - \frac{\gamma(\gamma+1)(\gamma+2)}{4!} m^{(4)}_{p,t+1} , \tag{42}$$

with

$$m^{(k)}_{p,t+1} = \mathrm{E}_t\left[(r_{p,t+1})^k\right] = \mathrm{E}_t\left[(\omega_t\, r_{sp,t+1} + (1-\omega)\, r_{f,t})^k\right] . \tag{43}$$

The statistical moments of the portfolio as specified by equation 43, are obtained by expanding the brackets, and noting that $\mathrm{E}_t\left[(r_{sp,t+1})^k\right] = m_{k,t+1}$, the non-standardized moment of the S&P 500.

Looking at equation 42, we find negative signs at the second and fourth moment term and a positive sign at the third moment term. Hence, expected utility increases when volatility declines, kurtosis declines, and skewness increases. Intuitively this makes sense since a higher spread in future returns (high volatility) as well as a higher probability of extreme returns (high kurtosis) is more risky. Similarly, there is less risk when there is a higher probability of (high) positive returns relative to (high) negative returns (positive skewness). Since we assume investors to be risk averse, utility should be lower when risk increases. This is consistent with the arguments developed by Scott and Horvath (1980) as well.

The second, third, and fourth moment of the S&P 500 are obtained from the different computational methods. The first moment, being the expected excess return of the S&P 500 over a one-month period, is fixed at 4.0 percent annually. This is equal to the value that Fama and French (2002) find, using a Dividend Discount Model to extract the expectation, over the period from 1872 to 1999. We leave this parameter fixed to isolate the effect of the information content in the second, third, and fourth moment. When forecasting the first moment, it can not be determined whether any outperformance is due to this forecasting power or due to the information content in the higher moments.

Finally, for the relative risk aversion parameter γ we use 4 as our base case scenario, the value that Bliss and Panigirtzoglou (2004) and Liu et al. (2007) find by using respectively calibration methods on the S&P 500 index and maximum likelihood estimation on the FTSE

100 index. To check the sensitivity of the results to γ, we also test for values equal to 2, 6, and 10.

Note that a positive first moment, in combination with a positive risk aversion parameter, implies that the optimal weight will never turn negative. It can become larger than one, in which case we borrow against the risk-free rate, r_f.

Performance Measures

The performance of the different trading strategies is measured along multiple dimensions. We use the Sharpe ratio (S), Sortino ratio (SR), and Treynor ratio (T), in addition to Jensen's alpha (α) and an alternative M^2 measure. The latter is alternative in the sense that we use downside risk in stead of the standard deviation. It can therefore be interpreted as the additional return of a strategy if is scaled to the same downside risk of the S&P 500. We use this measure as it essentially translates the Sortino ratio into a return, which is more intuitive in some instances. The measures are defined as

$$S = \frac{\overline{r}_p - \overline{r}_f}{\sigma_p} \ , \tag{44}$$

$$SR = \frac{\overline{r}_p - \overline{r}_f}{DR_p} \ , \tag{45}$$

$$T = \frac{\overline{r}_p - \overline{r}_f}{\beta_p} \ , \tag{46}$$

$$\alpha = \overline{r}_p - [\overline{r}_f + \beta_p(\overline{r}_{sp} - \overline{r}_f)] \ , \tag{47}$$

$$M^2 = (\overline{r}_p - \overline{r}_f)\frac{DR_{sp}}{DR_p} - (\overline{r}_{sp} - \overline{r}_f) \ , \tag{48}$$

with $\overline{r}_p$, σ_p, DR_p, and β_p the annualized mean excess return, standard deviation, downside risk, and beta (systematic risk) of the portfolio (strategy), and $\overline{r}_{sp}$ and DR_{sp} the annualized mean excess return and downside risk of the S&P 500. The definitions of DR_p and β_p are

$$DR_p = \sqrt{\frac{1}{N-1}\sum_{t=1}^{N}\min\{r_{e,p,t} - \overline{r}_{e,p,t}, 0\}^2}\ , \tag{49}$$

$$\beta_p = \frac{COV(r_{e,p,t}, r_{e,sp,t})}{V(r_{e,sp,t})}\ . \tag{50}$$

Note that the β is assumed constant over the sample-period, when computing the performance statistics. Actually this β equals the optimal weight in the S&P 500, and hence varies over time. However, when this variation in β would be taken into account, the Treynor ratio of all strategies will equal that of the S&P 500 and the alpha will be zero, per definition. By assuming a constant β, the 'timing ability' of the strategies will influence the value of the Treynor ratio and alpha.

The measure that is relevant depends on how the investment is used. When used as a stand-alone investment, the total risk and return is relevant. Hence, we have to look at either the Sharpe or Sortino ratio. When we add the trading strategy to a large portfolio of stocks, most of the unsystematic risk is diversified away. In such a setting the measure to consider depends on whether the investor considers the systematic risk relevant or not. Investors could not consider this risk relevant, as it can be neutralized, for example by going short β times the S&P 500 itself. In that case, the alpha is most relevant. When the investor does care about this systematic risk, the Treynor ratio is most relevant.

We mainly focus on the Sortino ratio and the alpha. The other measures are reported more for the sake of completeness. We prefer the Sortino ratio over the Sharpe ratio as it makes use of downside risk, in stead of the standard deviation. The advantage of downside risk, as opposed to the standard deviation, is that it does not assume risk to be symmetric and thereby also accounts for the higher moments in the return distribution.

We compare the performance measures of the five different trading strategies, to determine which computational method performs best. To measure the contribution of the higher moments, we compute the differences in performance measures as a result of the inclusion of a specific moment. Additionally, we determine what effect these moments have on the weights over time. Following a similar methodology as Jondeau and Rockinger (2008), this effect is measured by computing the distance, which we define as

$$Dist = \sum_{t=1}^{N} |\omega_t - \omega_t^*| \ , \tag{51}$$

with ω_t and ω_t^* respectively the weight in the S&P 500 of the strategy excluding and including the higher moment at month t.

5 Results

This section first discusses the summary statistics of the time series of moments, as derived from the different computational methods. Subsequently, the results of the static and rolling regressions are discussed. The results of the utility setting follow thereafter.

5.1 Time Series of Moments

In total we analyze 15 time series of standardized moments and 15 time series of non-standardized moments. Table 2 shows the summary statistics of the standardized moments, being the volatility, skewness, and kurtosis.

The mean of the option-implied volatility is approximately equal to the mean of the realized volatility. It is lower compared to the mean of the realized skewness and kurtosis. Only the finding regarding the skewness is expected, the the option-implied volatility and kurtosis are expected to be higher than their realized measure. This follows for example from the finding of Santa-Clara and Yan (2010) that risk implied by option-prices far exceeds realized risk. They postulate that this is due to option prices incorporating a risk premium for risks that do not always materialize. Given that a higher volatility and kurtosis as well as a lower skewness is more risky, the option-implied volatility, skewness, and kurtosis should respectively be higher, lower, and higher than their realized measures. This is consistent with the findings of Bollerslev et al. (2009), which also finds the option-implied volatility higher than the realized volatility.

The standard deviation of the option-implied series are found lower, higher, and lower than the realized series for the volatility, skewness, and kurtosis respectively. Furthermore, all first order AR terms are significantly different from zero. This indicates that an AR model is an appropriate model to use in forecasting the statistical moments.

The correlation coefficients between the moment variables, as used in the regressions, are displayed in Table 3. In general we find relatively low correlations between the log-volatility and skewness variables. The correlation between the log-volatility and log-kurtosis tend to be somewhat larger. The highest correlations show up between the skewness and log-kurtosis of the option-implied methods, which equal -0.93 and -0.89 respectively.

Table 2: Summary statistics of the volatility, skewness, and kurtosis series
This table reports the summary statistical of the volatility, skewness, and kurtosis series as computed by the methods as stated in the column headings. The statistics are all based on monthly data, however, the volatility series is annualized. The methods considered are the realized method, daily autoregressive method, using a estimation length of 315 days (AR^{315d}), monthly autoregressive method, using a estimation length of 15 months (AR^{15m}), model-free option-implied risk-neutral method, and Generalized Beta distribution of the second kind option-implied risk-neutral ($GB2^{Q}$) and real-world ($GB2^{P}$) method. All AR(1) coefficients are statistically different from zero, at a (two-sided) one percent significance level. The sample period extends from January 1996 to March 2010.

	Volatility			Skewness			Kurtosis		
	Mean	Std. dev.	AR(1)	Mean	Std. dev.	AR(1)	Mean	Std. dev.	AR(1)
Realized	0.20	0.12	0.80	0.09	0.29	0.19	10.00	3.38	0.44
AR^{315d}	0.22	0.15	0.78	0.03	0.08	0.25	9.89	2.91	0.71
AR^{15m}	0.22	0.15	0.83	0.06	0.15	0.22	10.39	3.49	0.56
Model-free	0.20	0.09	0.85	-1.38	0.37	0.58	6.35	1.87	0.55
$GB2^{Q}$	0.21	0.09	0.84	-1.18	0.35	0.42	5.93	1.05	0.52

Table 3: Correlation between moment variables
This table reports the Pearson correlation coefficients between the (transformed) moment series as used in the regressions. The variables that we consider are the log-volatility ($\ln$(vola)), skewness, and log-kurtosis ($\ln$(kurt)). All coefficients are based on monthly data. The sample period extends from January 1996 to March 2010.

		skewness	ln(kurt)
Realized	ln(vola)	-0.10	-0.31
	skewness		0.13
AR^{315d}	ln(vola)	-0.16	-0.41
	skewness		0.05
AR^{15m}	ln(vola)	-0.01	-0.39
	skewness		-0.02
Model-free	ln(vola)	0.13	-0.05
	skewness		-0.93
$GB2^{Q}$	ln(vola)	0.02	-0.36
	skewness		-0.89

5.2 Static Regressions

Table 4 shows the estimated coefficients of the static regression, with Newey-West t-statistics in parenthesis. The significance levels are based on two-sided p-values. Only the results of the moment variables are reported; the intercept and control variables are left out. The table also includes the F-statistic on the joint significance of all three variables.

The skewness coefficient should be interpreted as the additional one-month-ahead excess return (in monthly percentages) as a result of one-unit increase in the explanatory variable. The log-volatility and log-kurtosis coefficient should be interpreted as the additional one-month-ahead excess return (in monthly percentages) as a result of a 10 percent relative increase in the explanatory variable. For example, we would experience an additional excess return of 0.31 percent over the next month when the volatility variable as measured by the model-free method increases by 10 percent (for example from 20 to 22 percent).

Differences in coefficient estimates between regressing one moment variable versus all three becomes only present with the GB2^{Q} method. For this method, we find a volatility coefficient of 0.26 when estimated without the skewness and kurtosis, and 0.34 when estimated jointly (both at a 5 percent significance level). The coefficient of the regression using only the volatility variable is smaller, as the indirect effect of the kurtosis is negative due to the negative correlation of -0.89 between these two variables. Going forward, we focus on the jointly regressed variables.

We find significant volatility coefficients for the $\text{AR}^{315\text{d}}$, model-free, and GB2^{Q} method, with coefficients of 0.21, 0.31, and 0.34 respectively (at a 10, 5, and 5 percent significance level). None of the skewness and kurtosis coefficients are significant.

When the power curve pertaining to the relevant significance level is evaluated at the value of the estimated coefficient, the power is obtained as stated in the table[7]. The value of these figures range from 50 to 64 percent, which is not particularly high. Especially when both the (standardized) effect size and the sample size are relatively small, the null hypothesis becomes hard to reject. This might be the case with the skewness and log-kurtosis.

The F-statistics is significant for all methods and the R^2 measures are all relatively low.

[7] Notice that these figures depend on the estimated coefficient and the significance level. Therefore, caution should be taken when comparing these figures to each other. Section 7.2.1 in the Appendix shows the entire power curve for all, jointly regressed, significant log-volatility coefficients and the, jointly regressed, skewness and log-kurtosis coefficient of the model-free method. These curves are shown for a 5 and 10 percent significance level.

The latter is something we would expect given the noisiness of equity returns.

Our findings concerning the option-implied volatility coefficients is consistent with the work of Ammann et al. (2009). Our findings related to the realized volatility is consistent with the results of Amaya et al. (2011), who do not find any clear relation either. However, they also find a negative relation using realized skewness and a positive relation using realized kurtosis, whereas we do not find any significant relation for these two variables. The different methodology is expected to cause the differences in the results. Apart from that, they are likely to have a more powerful test as they use weekly returns and have therefore more observations at their disposal.

The non-significant skewness coefficients is a finding that is consistent with Conrad et al. (2009). However, they do find a significant relation using kurtosis, while we do not. Again, the different methodology is expected to cause the differences in the results. After all, they test for the intercept (alpha) in a Fama-French regressions. However, they also analyze a different sample period (1996 until 2005).

Table 4: Static regression results

This table reports the regression coefficients of the log-volatility (ln(vola)), skewness, and log-kurtosis (ln(kurt)) variable as used in the static regression, for the different computational methods. Additionally, it reports the F-statistic on the significance of all three variables combined, and the R^2. The computational methods that we consider are the realized method, daily autoregressive method, using a estimation length of 315 days (AR^{315d}), monthly autoregressive method, using a estimation length of 15 months (AR^{15m}), model-free option-implied risk-neutral method, and option-implied risk-neutral Generalized Beta Distribution of the Second Kind method ($GB2^{Q}$). The results are generated using the regression model as specified in equation 34 till 36. The intercept and coefficients of the control variables are not reported. The Newey-West t-statistics are stated in the parenthesis. The one, five, and ten percent (two-sided) significance level is indicated with three, two, and one stars. The power results from the evaluation of the power curve, pertaining to the significance level as stated in the table, at the value of the estimated coefficient. It relies on an asymptotic normality. The sample period extends from January 1996 to March 2010, so the number of observations included in the regression equals 176.

	ln(vola)				skewness				ln(kurt)						
	coef.	t-stat	sign.	power (%)	coef.	t-stat	sign.	power (%)	coef.	t-stat	sign.	power (%)	F-stat.	sign.	R^2 (%)
Realized	0.25	(1.58)													3.54
					0.62	(0.50)									1.30
									0.03	(0.21)					1.23
	0.26	(1.49)			0.48	(0.41)			0.07	(0.40)			4.48	***	3.78
AR^{315d}	0.22	(1.96)	**	50											2.80
					-2.06	(-0.49)									1.29
									-0.10	(-0.72)					1.39
	0.21	(1.69)	*	52	-1.42	(-0.33)			-0.02	(-0.10)			2.83	**	2.85
AR^{15m}	0.25	(1.61)													3.64
					2.06	(0.79)									1.51
									-0.05	(-0.46)					1.27
	0.25	(1.43)			1.89	(0.74)			0.01	(0.10)			4.71	***	3.91
Model-free	0.31	(1.92)	*	61											3.07
					0.49	(0.48)									1.31
									-0.06	(-0.45)					1.30
	0.31	(2.05)	**	54	1.18	(0.36)			0.09	(0.20)			3.49	**	3.22
$GB2^{Q}$	0.26	(1.98)	**	51											2.90
					-0.04	(-0.04)									1.20
									-0.09	(-0.41)					1.27
	0.34	(2.31)	**	64	2.21	(0.48)			0.28	(0.32)			3.38	**	3.16

5.3 Rolling Regressions

5.3.1 Parameter Variability

Figure 2 shows the estimated rolling coefficients of the moment variables over time. These estimates are generated using a window size of 40 months.

Fist of all, we find that all coefficients show substantial variation over time. Something Ammann et al. (2009) also find. For instance, the log-volatility coefficient of the model-free method attains values as low as -1.0 and as high as 2.3. Secondly, the skewness and kurtosis variable of all but the realized and AR^{15m} method pick up in value as off the year 2009, the start of the financial crisis. It seems as if the market perceives these higher moments as more relevant during financial market distress.

5.3.2 Forecasting Accuracy

In this section the results are discussed regarding the forecasting accuracy of the regression models. First, the contribution of the individual moments is discussed. Thereafter, the analysis follows when testing the different unrestricted models against each other.

Table 5 shows the Diebold-Mariano (DM) statistics of testing equal forecasting power between the restricted and unrestricted model, when using MSPE as loss function[8]. The figures are based on a MSPE difference derived by subtracting the MSPE of the unrestricted model from the MSPE of the restricted model. Since it involves nested models, a single-sided hypothesis is tested. The null and alternative hypothesis are respectively specified as $\mathrm{H_0}: \mathrm{MSPE_r} = \mathrm{MSPE_u}$ and $\mathrm{H_a}: \mathrm{MSPE_r} > \mathrm{MSPE_u}$, where r and u stands respectively for restricted and unrestricted. Another implication of testing nested models is that the distribution under the null does not converge to a standard normal distribution. Hence, simulated critical values are used, as reported by McCracken (1999). The variable(s) that are restricted are stated in the column headings of the table. We test for every moment individually as well as all three moments combined. The results are shown for different lengths of the estimation window.

Most significant statistics are found for the log-volatility variable, for a window length of

[8]The Diebold-Mariano statistic is defined as $\mathrm{DM} = \frac{\bar{d}}{\sqrt{\mathrm{V}(\hat{d})/P}}$, with $\bar{d} = \frac{1}{P}\sum_{t=R}^{T}(e^2_{r,t+1} - e^2_{u,t+1})$, where $e_{r,t+1}$ and $e_{u,t+1}$ are the forecasting errors of the restricted and unrestricted model respectively, at month t. This forecasting error is defined as $e_{t+1} = y_{t+1} - \widehat{y}_{t+1|T}$. R is the initial in-sample portion of the data, P, the number of one-step-ahead forecasts, and $T = R + P$.

20 months. Within this class, only the statistic of the AR^{315d} method is not significant. Still, there is no consistency in significance over different lengths of the estimation window. This also holds for the skewness and log-kurtosis variable.

Table 6 presents the DM statistics of testing equal forecasting power between the different unrestricted models, again using MSPE as loss function. The figures are based on a MSPE difference derived by subtracting the MSPE of the model as stated in the row headings, from the MSPE of the model as stated in the column headings. The null and alternative hypothesis are respectively specified as $H_0 : MSPE_r = MSPE_c$ and $H_a : MSPE_r \neq MSPE_c$, where r and c stands for row and column. Again, the results are shown for different lengths of the estimation window.

When solely looking at the signs, we find that the AR^{15m} method performs best, followed by the realized and model-free method, based on the number of models it beats in terms of accuracy. The $GB2^Q$ method performs worst in this respect. To a large extent, this contradicts the findings of the static regression. However, non of the statistics are significant (at a 10 percent level). Hence, we lack the statistical support to prefer one method over another.

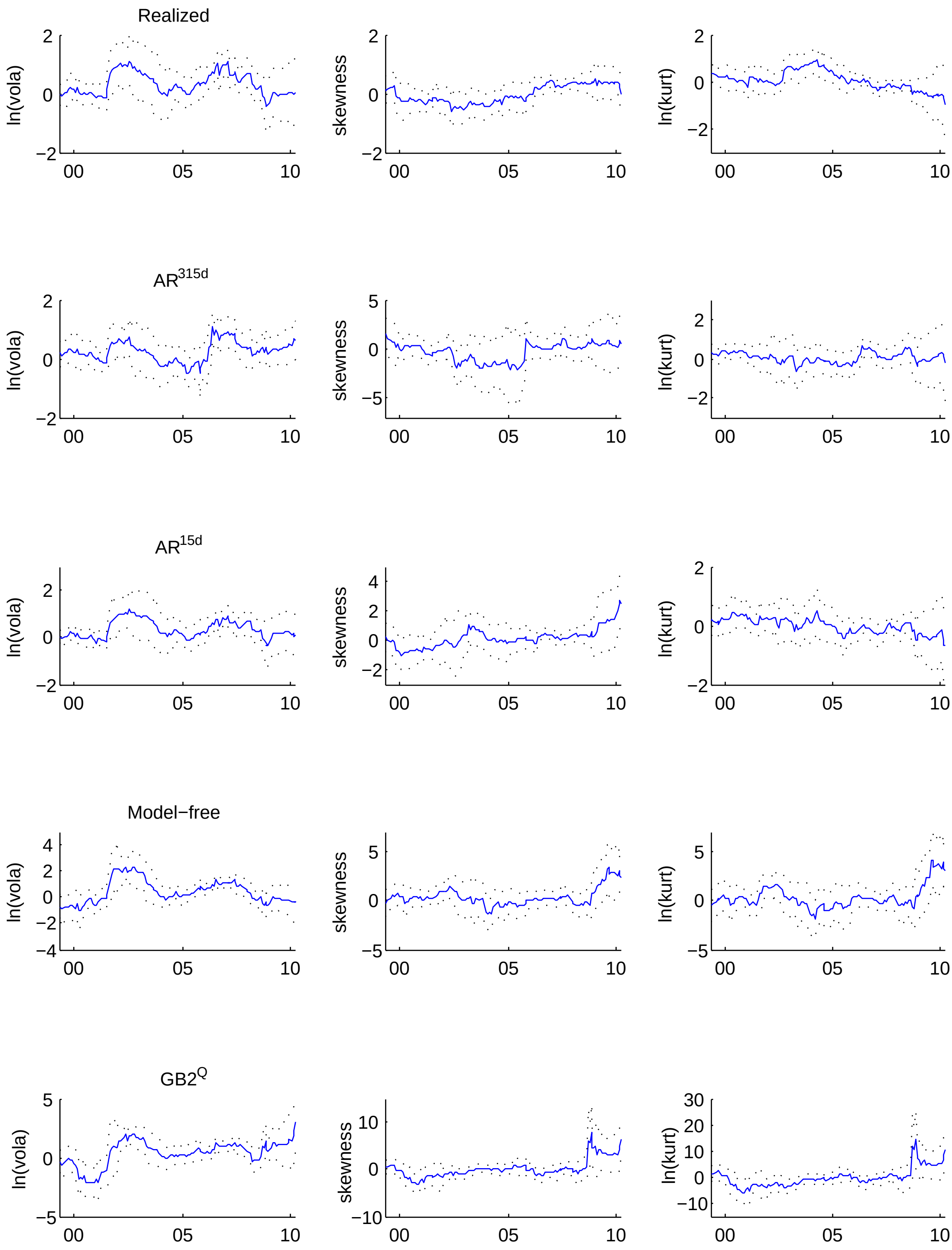

Figure 2: Rolling regression coefficients
This figure displays the rolling regression coefficients (in blue), with their 95 percent confidence bounds (in dotted black) of all moments, for a window length of 40 months. Newey-West standard errors are used in constructing the confidence bounds. The sample period extends from January 1996 to March 2010 (including estimation period).

Table 5: Diebold-Mariano statistics of nested models
This table exhibits the DM statistics of testing equal forecasting power, between the restricted and unrestricted model, using MSPE as loss function. The figures are based on a MSPE difference derived by subtracting the MSPE of the unrestricted model from the MSPE of the restricted model. We test for the log-volatility (ln(vola)), skewness, log-kurtosis (ln(kurt)), and all three variables combined, as stated in the columns. We use a rolling window regression, with a window length equal to 20, 40, and 60 months. The one, five, and ten percent (one-sided) significance level is indicated with three, two, and one stars. The critical values are obtained by linearly interpolating the two closest surrounding simulated values as reported by McCracken (1999). The sample period extends from January 1996 to March 2010.

		ln(vola)		skewness		ln(kurt)		All	
		DM	sign.	DM	sign.	DM	sign.	DM	sign.
20	Realized	0.17	**	-1.32		-1.74		-2.33	
	AR^{315d}	-1.43		-1.13		-2.13		-1.60	
	AR^{15m}	0.03	**	-0.80		-1.74		-1.52	
	Model-free	0.07	**	-1.42		-1.27		-1.36	*
	$GB2^{Q}$	0.16	**	-1.86		-1.60		-2.02	
40	Realized	-0.69		-1.73		-1.32		-2.18	
	AR^{315d}	-0.84		-1.55		-2.15		-1.94	
	AR^{15m}	-0.46		-1.31		-1.57		-1.69	
	Model-free	0.11	*	-1.85		-1.33		-1.23	
	$GB2^{Q}$	-0.22		-0.62		-0.79		-0.92	
60	Realized	-0.01		-1.94		0.08		-0.74	
	AR^{315d}	-0.57		-1.68		-1.36		-1.37	
	AR^{15m}	0.08		0.12	*	-0.28		0.01	**
	Model-free	-0.14		-1.12		-0.58		-0.58	
	$GB2^{Q}$	-0.38		-1.10		-0.76		-1.14	

Table 6: Diebold-Mariano statistics of non-nested models
This table exhibits the DM statistics of testing equal forecasting power, between the different unrestricted models, using MSPE as loss function. It is constructed by subtracting the MSPE of the model as stated in the row headings from the MSPE of the model as stated in the column headings. We use a rolling window regression, with an window length equal to 20, 40, and 60 months. The one, five, and ten percent (two-sided) significance level is indicated with three, two, and one stars. The sample period extends from January 1996 to March 2010.

		Realized		AR^{315d}		AR^{15m}		Model-free		$GB2^{Q}$	
		DM	sign.	DM	sign.	DM	sign.	DM	sign.	DM	sign.
20	Realized	-		1.35		-0.31		0.01		0.79	
	AR^{315d}	-1.35		-		-1.24		-1.03		-0.71	
	AR^{15m}	0.31		1.24		-		0.25		1.25	
	Model-Free	-0.01		1.03		-0.25		-		0.87	
	$GB2^{Q}$	-0.79		0.71		-1.25		-0.87		-	
40	Realized	-		-0.52		-1.03		-0.43		0.06	
	AR^{315d}	0.52		-		-0.47		-0.14		0.22	
	AR^{15m}	1.03		0.47		-		0.13		0.40	
	Model-Free	0.43		0.14		-0.13		-		0.37	
	$GB2^{Q}$	-0.06		-0.22		-0.40		-0.37			
60	Realized	-		0.61		-1.02		0.01		0.57	
	AR^{315d}	-0.61		-		-1.12		-0.43		0.08	
	AR^{15m}	1.02		1.12		-		0.66		1.19	
	Model-Free	-0.01		0.43		-0.66		-		0.92	
	$GB2^{Q}$	-0.57		-0.08		-1.19		-0.92		-	

5.4 Utility Framework

5.4.1 Performance Trading Strategies

This section contains the analysis of the performance of the different trading strategies, in the base case scenario. Starting with the optimal portfolio weights, Figure 3 shows these for the model-free method over time. Clearly, the optimal portfolio weight is high when volatility is low and vice versa. This general pattern holds for all methods. The mean value of these weights, as well as the maximum and standard deviation, are higher for the realized and AR method than for the option implied methods. This is consistent with the summary statistics of the moment time series themselves.

Table 7 reports the performance measures of the different trading strategies. As shown, all strategies experienced a lower excess return than the S&P 500. However, this also holds for both the total and systematic risk. The latter is partly the result of the relatively low correlation with the S&P 500. These correlations all fall in the range between 0.60 to 0.76.

Looking across all risk adjusted performance measures, we find both the $\text{AR}^{315\text{d}}$ and model-free method to perform best, followed by the GB2^{Q}, realized and $\text{AR}^{15\text{m}}$ method. Remarkably, the $\text{AR}^{315\text{d}}$ method performs much better than the $\text{AR}^{15\text{m}}$ method. This is something we would not expect given that the MSPE of the monthly AR model is slightly lower than the daily AR model. Apparently, there is some information in the daily observations of the realized monthly moments that is not captured by the monthly model.

Zooming in on the Sortino ratio, both the $\text{AR}^{315\text{d}}$ and model-free method perform equally well with a ratio of 0.30, followed by the GB2^{Q}, realized, and $\text{AR}^{15\text{m}}$ method, with ratios equal to 0.26, 0.19, and 0.18, respectively. Comparing these to a ratio of 0.24 for the S&P 500, we find that only the realized and $\text{AR}^{15\text{m}}$ method perform worse than the market. The Sharpe ratio shows a similar picture.

The M^2 measure equals 1.27, 1.21, 0.34, -1.12, and -1.28 percent for the model-free, $\text{AR}^{315\text{d}}$, GB2^{Q}, realized, and $\text{AR}^{15\text{m}}$ method respectively. It shows that that the Sortino ratio of the model-free method is actually slightly higher than the $\text{AR}^{315\text{d}}$, however, this difference is only marginal. Applying the law of conditional expectations, results in a probability of performing better than the S&P 500 on a certain day that is larger than 50 percent for all methods, and particularly large for the $\text{AR}^{315\text{d}}$ method. However, the average daily outperformance is generally smaller than the average underperformance.

Measured in alpha, all methods outperform the S&P 500. The alphas equal 1.02, 0.65, 0.43, 0.29, and 0.28 percent for the $\text{AR}^{315\text{d}}$, model-free, GB2^{Q}, realized, and $\text{AR}^{15\text{m}}$ method respectively. The Treynor ratio shows a similar picture, but with the realized and $\text{AR}^{15\text{m}}$ method switching positions.

The results are consistent with those of Jondeau and Rockinger (2008), in the sense that we also find a positive economic value as a result of distributional timing. To determine this value, we subtract the performance of the realize method from that of the $\text{AR}^{315\text{d}}$ method. This yields an annual M^2 measure and alpha of 2.33 and 0.73 percent annually. The 1.40 percent that Jondeau and Rockinger (2008) find, falls within this range. Our finding concerning the added value of the higher moments is not in accordance with their work.

5.4.2 Consistency Over Time

Figures 4, 5, and 6 show respectively the cumulative excess return, scaled cumulative excess return (to the same downside risk of the S&P 500), and the M^2 measure, over time. Especially from the latter it becomes clear that none of the methods perform consistently over time. The positive M^2 measure of all methods is concentrated at the last three and a half years of our sample period. Halfway 2006, all methods yield a negative M^2 measure.

Apart from the last three and a half years, the option-implied methods also performed well from 1996 to 1999. However, in the four subsequent years the M^2 measure declined again. Moreover, the realized method barely experiences periods of positive outperformance.

5.4.3 Contribution of Higher Moments

Table 8 shows the difference in the various performance measures as caused by the third and fourth moment, in Panel A and Panel B respectively. These figures are computed by subtracting the values without from the values with the specific higher moment. Hence, positive figures indicate a improvement as a result of the inclusion. This table also contains the distance measure.

The distance in weights caused by the third and fourth moment is largest using the option-implied methods, in particular the model-free method. The distances caused by the fourth moment are smaller than those caused by the third moment for the realized and option-implied methods. This is also what we would expect given that in a Taylor expansion the impact of higher orders diminishes. The mean and standard deviation of the weights only change for the third moment of the option-implied methods. These effects are all equal to -0.02. Hence, negative but small.

The Sortino and Sharpe ratio only change for the third moment of the option-implied methods. This is in line with their higher distance measure. The M^2 measure shows larger deviations than the Sortino ratio as this measure is more sensitive due to scaling. Some of the differences are positive, like the Sharpe ratio of the option-implied method, which increases by 0.01 when including the third moment. However, most figures are negative. This suggests that in general the inclusion of these moments only reduce the performance. Nonetheless, we do not consider these contributions to be economically significant.

These findings are consistent with the results of the static regression, where we do not find significant coefficients for the third and fourth moment either (these are standardized, however).

5.4.4 Restricted Weights

In this section we impose an upper bound restriction, c, on ω. We test for different values of c, being 1.6, 1.3, 1.0, 0.7, and 0.5. Given that the maximum value of ω equals 1.08 and 1.10 for the model-free and GB2^{Q} method respectively, the first restriction is not effective for these methods. It is effective for the remaining methods, knowing that these maxima are 1.77, 1.67, and 1.63 for the realized, AR^{315d}, and AR^{15m} method respectively. Table 9 exhibits the results.

We find for all methods a gradually improving performance when the restriction becomes more stringent, up till a tipping point, after which the performance deteriorates. The optimal restriction differs per method and often per performance measure as well.

For the realized method, the optimal restriction equals 1.0 for the Sortino and Sharpe ratio, and 1.6 for the alpha and Treynor ratio. The AR^{315d} method attains the best performance at a restriction of 1.0 for all measures. For the AR^{15m} method, the optimal point is reached at a restriction of 0.7 for the Sortino and Sharpe ratio, and 1.0 for the alpha and Treynor ratio. Finally, both option-implied methods perform best when imposing a restriction of 0.7, for all performance measures.

In terms of the Sortino ratio, we find maximum improvements of 0.01, 0.04, 0.02, 0.02, and 0.04 for the realized, $\text{AR}^{315\text{d}}$, $\text{AR}^{15\text{m}}$, model-free, and GB2^{Q} method respectively. The alphas improve respectively by a maximum of 0.04, 0.11, 0.02, 0.00, and 0.10 percent. Hence, the improvements are small.

5.4.5 Sensitivity Analysis

Since the optimal portfolio weights depend directly on the risk aversion parameter, we check if the results are robust for changes in this variable. Table 10 provides the results for values of γ equal to 2, 4, 6, and 10.

For all methods, both the mean and standard deviation of the optimal weight shrink when γ increases. This makes sense, as higher risk aversion results in a lower preference to invest in the risky asset. The Sharpe, Sortino, and Treynor ratio are all remarkably stable over the different values of γ. Take for example the Sortino ratio of the model-free method which remains constant at a value of 0.30. The exception is the $\text{AR}^{315\text{d}}$ method, for γ equal to two. In that case the performance ratios are substantially higher. For example, the Sortino ratio jumps to 0.34, compared to a value of 0.30 for the other values of γ.

The alphas decline considerably as γ increases. For example, the model-free method yields alphas of 1.25, 0.65, 0.44, and 0.26 percent for γ equal to 2, 4, 6, and 10 respectively. For all methods this measure declines 79 to 82 percent when γ moves from 2 to 10. This decline is due to the fact that the alphas are subject to scaling. To see this, consider the definition of alpha in regression form $r_{p,t} - r_{f,t} = \alpha + \beta_p(r_{sp,t} - r_{f,t}) + \varepsilon_t$. When scaling the excess return of the strategy (portfolio) with a scaler, c, we get $c\,(r_{p,t} - r_{f,t}) = c\,\alpha + c\,\beta_p(r_{sp,t} - r_{f,t}) + c\,\varepsilon_t$. Therefore, the alpha scales by the same amount as the excess return. Increasing γ effectively scales down the portfolio weight and hence the portfolio return. This explanation fits the data pretty well, since the mean value of the weights decline by 78 to 79 percent, for all methods, when γ increases from 2 to 10, which is rather close to the decline in alphas by 79 to 82 percent. Hence, the alpha can be a misleading measure in this case.

We conclude that the sensitivity of the performance to variation in the relative risk aversion parameter γ is low.

5.4.6 Multiple Estimation Lengths with Autoregressive Models

When calibrating the AR models, smaller lengths of the estimation window seem to improve the accuracy of the daily model, while larger lengths are more favorable for the monthly model. The differences are small, however. Since there is not one window length that stands out, we also test the AR models estimated on different lengths than 15 months (or 315 trading-days). Lengths that we consider are 10, 15, 20, and 25 months, equalling 210, 315, 420, and 525 trading-days.

Table 11 exhibits the results of the different AR models. We find that the daily models all perform better than the monthly models, expressed in the all performance measures. Measured in Sortino and Sharpe ratio, all daily models outperform the S&P 500 while all monthly models underperform the S&P 500. When looking at the alpha and Treynor ratio, also the monthly models perform better than the S&P 500.

When the estimation window increases for the daily model, the performance gradually improves. For the monthly model, the performance becomes slightly better when the estimation window gets shorter. This is not expected since the MSPE becomes larger (smaller) for the daily (monthly) model, when the estimation window increases.

The best daily and monthly models, yield Sortino ratios of 0.32 and 0.19 for the AR^{525d} and AR^{10m} model respectively. In terms of alpha, these models yield respectively 1.13 and 0.33 percent. Comparing this to the performance of the model-free method, the AR^{525d} model performs better across all performance measures.

Table 7: Performance statistics

This Table displays the performance statistics of the five different trading strategies. The strategies are created by maximizing the expected utility each month. All figures are annualized ratios or annualized percentages, with the exception of the two conditional expectations, which are daily percentages. The results are based on a relative risk aversion parameter of 4. The sample period extends from January 1996 to March 2010.

	Realized	AR^{315d}	AR^{15m}	Model-free	$GB2^{Q}$	S&P500
Mean weight	0.47	0.42	0.41	0.34	0.32	1.00
Std. dev. weight	0.40	0.38	0.38	0.23	0.23	0.00
Min weight	0.01	0.01	0.01	0.02	0.02	1.00
Max weight	1.77	1.68	1.63	1.08	1.10	1.00
Excess return	1.45	2.08	1.25	1.60	1.31	5.00
Std deviation	7.39	6.76	6.62	5.10	4.87	20.53
Downside risk	7.74	6.92	6.94	5.26	5.06	20.65
Correlation	0.65	0.65	0.60	0.76	0.74	1.00
Beta	0.23	0.21	0.19	0.19	0.18	1.00
Sharpe ratio	0.20	0.31	0.19	0.31	0.27	0.24
Sortino ratio	0.19	0.30	0.18	0.30	0.26	0.24
Treynor ratio (x10)	0.62	0.98	0.64	0.84	0.75	0.50
Alpha	0.29	1.02	0.28	0.65	0.43	0.00
M^2 measure	-1.12	1.21	-1.28	1.27	0.34	0.00
P($M^2 > 0$)	51.20	52.02	50.69	51.72	51.31	0.00
E[$M^2 \| M^2 > 0$]	0.60	0.62	0.65	0.47	0.50	0.00
E[$M^2 \| M^2 < 0$]	-0.64	-0.67	-0.68	-0.49	-0.52	0.00

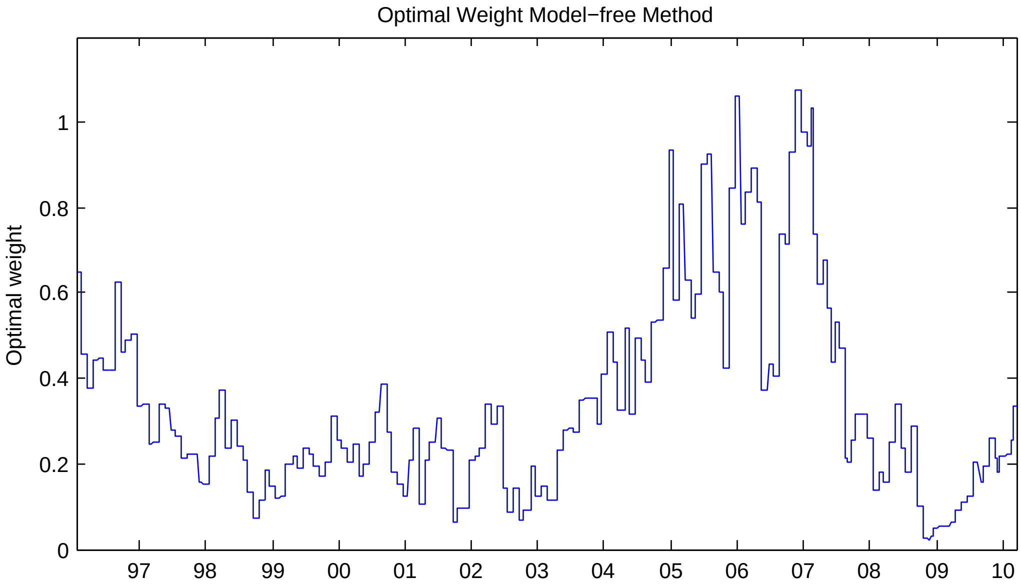

Figure 3: Optimal weight of the model-free trading strategy
This figure displays the optimal weight in the S&P 500 over time, when maximizing expected utility, for the model-free trading strategy. It is based on the second, third, and fourth moment, for a relative risk aversion parameter of 4. The sample period extends from January 1996 to March 2010.

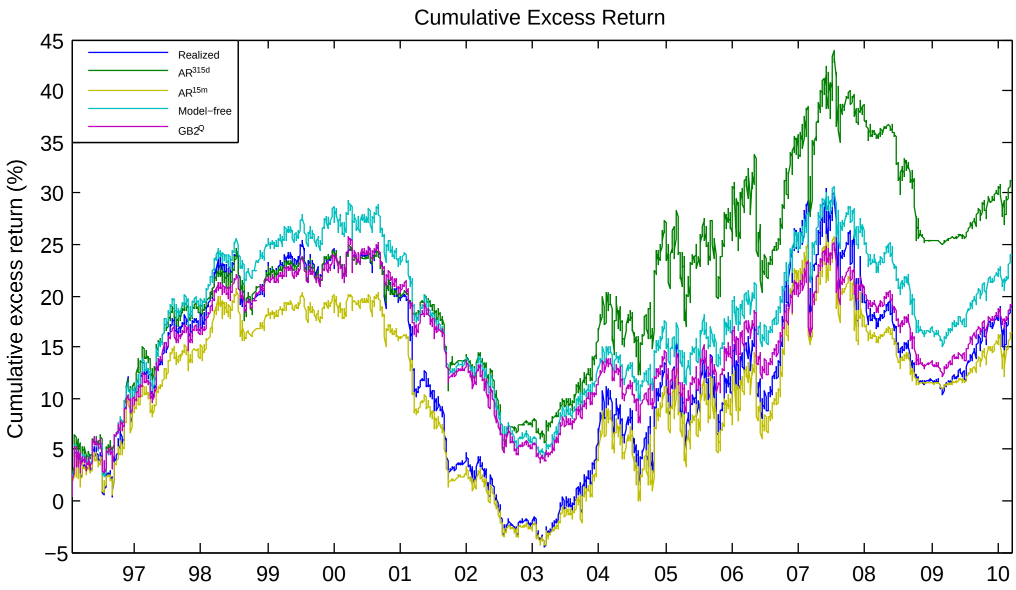

Figure 4: Cumulative excess return of trading strategies
This figure displays the cumulative excess return of the five different trading strategies over time. The sample period extends from January 1996 to March 2010.

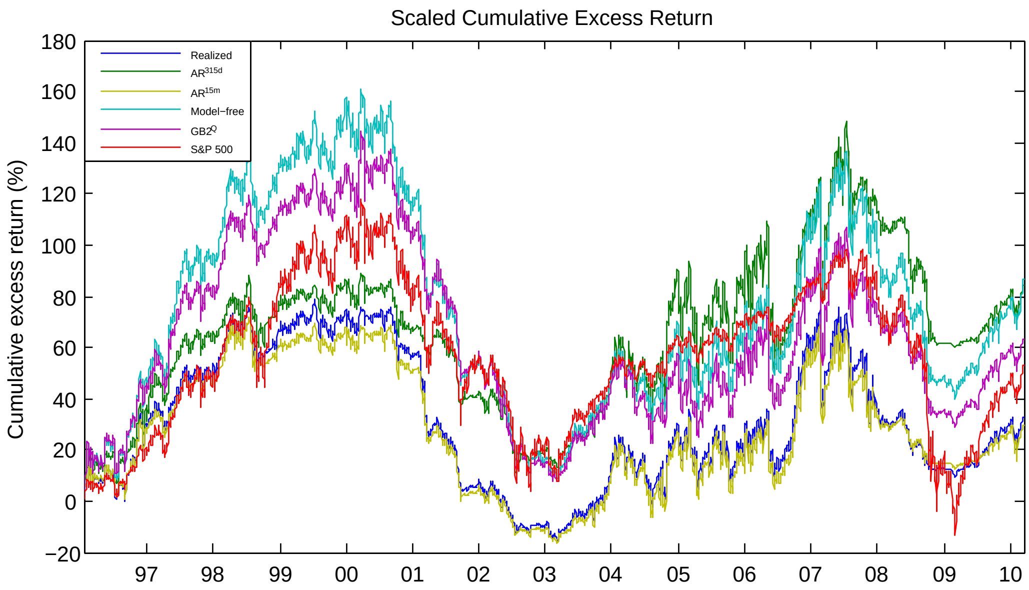

Figure 5: Scaled cumulative excess return of trading strategies
This figure displays the cumulative excess return of the five different trading strategies over time, when scaled to the downside risk of the S&P 500, and of the the S&P 500 itself. The sample period extends from January 1996 to March 2010.

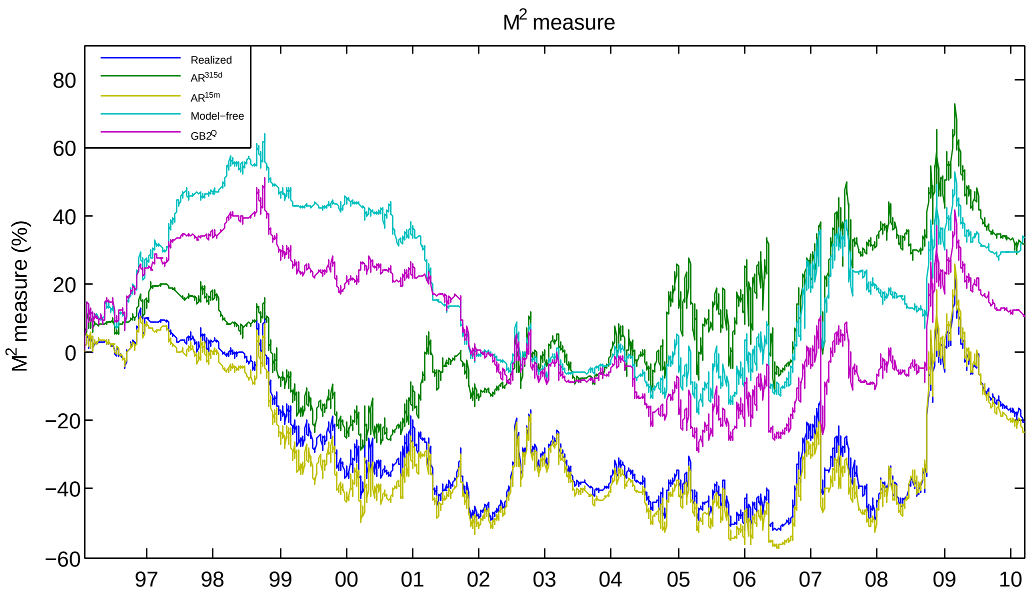

Figure 6: M^2 measure of trading strategies
This figure displays the M^2 measure of the five different trading strategies over time. The sample period extends from January 1996 to March 2010.

Table 8: Performance contribution of higher moments
This table displays the performance contribution, as well as the distance in weights, caused by the third and fourth moment, in Panel A and Panel B respectively. The figures are constructed as the performance based on a strategy with the specific higher moment minus the performance of the strategy without the specific higher moment. All figures are differences of annualized ratios or annualized percentages. The results are based on a relative risk aversion parameter of 4. The sample period extends from January 1996 to March 2010.

Panel A: $\mathbf{m}_3$	Realized	AR^{315d}	AR^{15m}	Model-free	$GB2^{Q}$
Distance	12.38	3.28	1.86	75.53	67.53
Δ Mean weight	0.00	0.00	0.00	-0.02	-0.02
Δ Std. dev. weight	0.00	0.00	0.00	-0.02	-0.02
Δ Sharpe ratio	0.00	0.00	0.00	0.01	0.01
Δ Sortino ratio	0.00	0.00	0.00	-0.01	-0.01
Δ Treynor ratio (x10)	0.00	0.00	0.00	0.00	0.00
Δ Alpha	0.00	0.00	0.00	0.00	0.00
Δ M^2 measure	0.01	0.00	0.00	-0.06	-0.06
Panel B: $\mathbf{m}_4$	Realized	AR^{315d}	AR^{15m}	Model-free	$GB2^{Q}$
Distance	5.71	7.30	7.18	13.61	12.01
Δ Mean weight	0.00	0.00	0.00	0.00	0.00
Δ Std weight	0.00	0.00	0.00	0.00	0.00
Δ Sharpe ratio	0.00	0.00	0.00	0.00	0.00
Δ Sortino ratio	0.00	0.00	0.00	0.00	0.00
Δ Treynor ratio (x10)	0.00	0.00	0.00	0.00	0.00
Δ Alpha	0.00	0.00	0.00	0.00	0.00
Δ M^2 measure	0.00	-0.01	0.00	-0.01	-0.01

Table 9: Performance when restricting ω
This table shows the performance statistics of the five different trading strategies, when restricting the optimal weight, ω, with an upper bound constraint, c, as $\omega \leq c$. We use upper bound constraints equal to 1.6, 1.3, 1.0, 0.7, and 0.5. All figures are differences of annualized ratios or annualized percentages. The results are based on a relative risk aversion parameter of 4. The sample period extends from January 1996 to March 2010.

	c	Realized	AR^{315d}	AR^{15m}	Model-free	$GB2^{Q}$
	1.6	0.00	0.00	0.00	-	-
	1.3	-0.01	0.00	0.00	-	-
Δ Mean weight	1.0	-0.04	-0.03	-0.03	0.00	0.00
	0.7	-0.09	-0.08	-0.08	-0.02	-0.02
	0.5	-0.15	-0.13	-0.13	-0.05	-0.04
	1.6	0.00	0.00	0.00	-	-
	1.3	-0.02	-0.01	-0.01	-	-
Δ Std. dev. weight	1.0	-0.08	-0.07	-0.06	0.00	0.00
	0.7	-0.16	-0.14	-0.14	-0.04	-0.04
	0.5	-0.24	-0.21	-0.21	-0.09	-0.08
	1.6	0.01	0.00	0.00	-	-
	1.3	0.01	0.01	0.00	-	-
Δ Sharpe ratio	1.0	0.01	0.04	0.01	0.00	0.01
	0.7	0.01	0.02	0.02	0.01	0.03
	0.5	-0.02	-0.01	-0.01	-0.01	0.02
	1.6	0.01	0.00	0.00	-	-
	1.3	0.01	0.01	0.01	-	-
Δ Sortino ratio	1.0	0.01	0.04	0.02	0.00	0.01
	0.7	0.01	0.02	0.02	0.02	0.04
	0.5	-0.02	-0.01	-0.01	-0.01	0.03
	1.6	0.02	-0.00	0.00	-	-
	1.3	0.01	0.02	0.01	-	-
Δ Treynor ratio (x10)	1.0	0.01	0.07	0.02	-0.00	0.01
	0.7	-0.05	-0.06	-0.02	0.01	0.06
	0.5	-0.16	-0.20	-0.13	-0.09	0.01
	1.6	0.04	-0.00	0.01	-	-
	1.3	0.01	0.03	0.01	-	-
Δ Alpha	1.0	0.00	0.11	0.02	-0.00	0.03
	0.7	-0.13	-0.21	-0.05	0.00	0.10
	0.5	-0.36	-0.53	-0.26	-0.19	-0.01
	1.6	0.14	0.01	0.03		
	1.3	0.13	0.19	0.12	-	-
Δ M^2 measure	1.0	0.28	0.84	0.31	0.01	0.14
	0.7	0.14	0.45	0.36	0.31	0.75
	0.5	-0.43	-0.24	-0.19	-0.23	0.55

Table 10: Performance for different values of γ
This table reports the performance statistics of the five different trading strategies, for different values of the relative risk aversion parameter γ. We use values of γ equal to 2, 4, 6, and 10. All figures are annualized ratios or annualized percentages. The results are based on a relative risk aversion parameter of 4 The sample period extends from January 1996 to March 2010.

	γ	Realized	AR^{315d}	AR^{15m}	Model-free	$GB2^{Q}$
Mean weight	2	0.86	0.78	0.77	0.67	0.62
	4	0.47	0.42	0.41	0.34	0.32
	6	0.31	0.28	0.28	0.23	0.21
	10	0.19	0.17	0.17	0.14	0.13
Std. dev. weight	2	0.64	0.65	0.65	0.45	0.45
	4	0.40	0.38	0.38	0.23	0.23
	6	0.27	0.26	0.26	0.16	0.16
	10	0.16	0.15	0.15	0.10	0.09
Sharpe ratio	2	0.21	0.35	0.20	0.31	0.27
	4	0.20	0.31	0.19	0.31	0.27
	6	0.20	0.31	0.19	0.31	0.27
	10	0.20	0.31	0.19	0.31	0.27
Sortino ratio	2	0.20	0.34	0.19	0.30	0.26
	4	0.19	0.30	0.18	0.30	0.26
	6	0.19	0.30	0.18	0.30	0.26
	10	0.19	0.30	0.18	0.30	0.26
Treynor ratio (x10)	2	0.63	1.05	0.66	0.83	0.75
	4	0.62	0.98	0.64	0.84	0.75
	6	0.62	0.98	0.64	0.84	0.75
	10	0.62	0.98	0.64	0.85	0.75
Alpha	2	0.58	2.24	0.58	1.25	0.87
	4	0.29	1.02	0.28	0.65	0.43
	6	0.19	0.68	0.19	0.44	0.29
	10	0.12	0.41	0.11	0.26	0.18
M^2 measure	2	-0.85	2.03	-0.98	1.24	0.42
	4	-1.12	1.21	-1.28	1.27	0.34
	6	-1.12	1.21	-1.28	1.28	0.35
	10	-1.12	1.21	-1.28	1.28	0.35

Table 11: Performance of AR models
This table reports the performance statistics of the trading strategies based on different autoregressive (AR) models. These models are fitted on either daily or monthly data, and vary in the length of the estimation window. This length ranges from from 10 to 25 months (or 210 to 525 trading-days). The strategies are created by maximizing the expected utility each month. All figures are annualized ratios or annualized percentages. The results are based on a relative risk aversion parameter of 4. The sample period extends from January 1996 to March 2010.

	AR^{210d}	AR^{315d}	AR^{420d}	AR^{525d}	AR^{10m}	AR^{15m}	AR^{20m}	AR^{25m}
Mean weight	0.43	0.42	0.40	0.40	0.43	0.41	0.40	0.40
Std. dev. weight	0.39	0.38	0.37	0.37	0.39	0.38	0.38	0.38
Min weight	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Max weight	1.59	1.68	1.64	1.54	1.82	1.63	1.71	1.65
Excess return	1.83	2.08	2.08	2.15	1.34	1.25	1.14	1.17
Std deviation	6.93	6.76	6.55	6.48	6.81	6.62	6.50	6.42
Downside risk	7.17	6.92	6.70	6.61	7.16	6.94	6.83	6.73
Correlation	0.65	0.65	0.65	0.64	0.61	0.60	0.60	0.60
Beta	0.22	0.21	0.21	0.20	0.20	0.19	0.19	0.19
Sharpe ratio	0.26	0.31	0.32	0.33	0.20	0.19	0.18	0.18
Sortino ratio	0.25	0.30	0.31	0.32	0.19	0.18	0.17	0.17
Treynor ratio (x10)	0.83	0.98	1.00	1.06	0.66	0.64	0.60	0.62
Alpha	0.73	1.02	1.04	1.13	0.33	0.28	0.19	0.23
M^2 measure	0.27	1.21	1.40	1.71	-1.14	-1.28	-1.54	-1.42
P($M^2 > 0$)	51.56	52.02	51.64	51.83	50.91	50.69	50.66	50.55
E[$M^2 \| M^2 > 0$]	0.61	0.62	0.62	0.62	0.64	0.65	0.65	0.65
E[$M^2 \| M^2 < 0$]	-0.65	-0.67	-0.65	-0.65	-0.67	-0.68	-0.68	-0.68

6 Conclusion

Using static regressions, we find a significant volatility coefficient for all but the realized and AR^{15m} methods. The coefficients are as such that a 10 percent relative increase in volatility results in a additional excess return, one month ahead, of 0.21, 0.31, and 0.34 percent, for the AR^{315d}, model-free, and $GB2^{Q}$ method respectively. The skewness and kurtosis coefficients are not found significant for any method.

The rolling window regressions show that the coefficients vary substantially over time. Furthermore, we find some support for the volatility variable in improving the accuracy of the forecasts. When comparing the accuracy of the different methods, we lack the statistical support to prefer one model over another.

Regarding the utility framework, we find that the model-free and AR^{315d} method perform best with Sortino ratio's of 0.30, compared to a ratio of 0.24 for the S&P 500. Second best is the $GB2^{Q}$ method, followed by the realized, and AR^{15m} method, with ratios equal to 0.26, 0.19, and 0.18 respectively. Hence, only the realized and AR^{15m} method perform worse than the market. Measured in alpha, all methods outperform the S&P 500. These alphas equal 1.02, 0.65, 0.43, 0.29, and 0.28 percent for the AR^{315d}, model free, $GB2^{Q}$, realized, and AR^{15m} method respectively. Effectively all of this outperformance is attributable to the second moment; the contribution of the third and fourth moment are economically not significant.

Non of the strategies perform consistently over time. Measured halfway 2006, the M^2 measure of all strategies was negative. Hence, the positive overall M^2 measure is due to the last three and a half years of our sample period. Measured over different values of the relative risk aversion parameter, the performance is stable.

The performance of all strategies improve moderately when imposing an upper bound restriction on the optimal weight. When the restriction gets too stringent, the performance deteriorates. Looking at the Sortino ratio we find maximum improvements of 0.01, 0.04, 0.02, 0.02, and 0.04 for the realized, AR^{315d}, AR^{15m}, model-free, and $GB2^{Q}$ methods respectively. The alphas improve respectively by a maximum of 0.04, 0.11, 0.02, 0.00, and 0.10 percent annually.

Testing for different estimation lengths for the AR models, we find that the performance of the daily model gradually improves when the estimation window increases. For the monthly model, the performance becomes slightly better when the estimation window gets shorter.

Overall, the daily model performs better than the monthly model. The best performing AR model yields a Sortino ratio and alpha of respectively 0.32 and 1.13 percent, thereby outperforming the model-free method on both measures.

We conclude that the volatility variable as derived from the daily AR method contains most information on future returns, followed by the model-free method. We do not find any indication that the higher moments contain information on future returns. Although the economic performance of the AR^{315d} and model-free method are equal, we still classify the AR method as best since its performance improves when increasing the estimation window[9].

Several extensions to this research may be considered. Fist of all, it would be interesting to analyze if there is any performance improvement when the risk aversion parameter, γ, is varied over time, in stead of kept constant. For example, the value of this parameter could be determined each month again, using either calibration methods or maximum likelihood estimation, like respectively Bliss and Panigirtzoglou (2004) and Liu et al. (2007) do. Alternatively, this parameter can be related directly to the option-implied market price of risk.

Secondly, extending the utility framework with a forecasting method for the first moment of the S&P 500 may also yield better performance. With the results presented in this paper, the additional performance improvement can be determined as a result of this modification. The regression model based on the volatility variable of the AR or model-free method could for example be used for this purpose.

Finally, it would be interesting to see how the different methods perform on alternative equity indices, or even fixed income indices and for different lengths of the holding period.

[9]The performance improves even more when using the alternative measure of the realized moments. See Section 7.2.2 in the Appendix.

References

D. Amaya, P. Christoffersen, K. Jacobs, and A. Vasquez. Does realized skewness and kurtosis predict the cross-section of equity returns? Technical report, Working paper, University of Houston, 2011.

M. Ammann, M. Verhofen, and S. Suss. Do implied volatilities predict stock returns. *Journal of Asset Management*, 10(4):222–234, 2009.

T.G. Andersena, T. Bollerslev, F.X. Dieboldc, and P. Labysd. Realized volatility and correlation. 1999.

G. Bakshi, N. Kapadia, and D. Madan. Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies*, 16(1):101, 2003.

O.E. Barndorff-Nielsen and N. Shephard. Realized power variation and stochastic volatility models. *Bernoulli*, pages 243–265, 2003.

D.S. Bates. The crash of '87: Was it expected? the evidence from options markets. *Journal of Finance*, pages 1009–1044, 1991.

M.W. Baxter and A.J.O. Rennie. Financial calculus: An introduction to derivative pricing. *Cambridge University Press*, 2008.

R.R. Bliss and N. Panigirtzoglou. Option-implied risk aversion estimates. *The Journal of Finance*, 59(1):407–446, 2004.

T. Bollerslev, G. Tauchen, and H. Zhou. Expected stock returns and variance risk premia. *Review of Financial Studies*, 22(11):4463, 2009.

R.M. Bookstaber and J.B. McDonald. A general distribution for describing security price returns. *Journal of Business*, pages 401–424, 1987.

R.M. Bookstaber and J.B. McDonald. Option pricing for generalized distributions. *Communications in Statistics*, pages 4053–4068, 1991.

M.M. Carhart. On persistence in mutual fund performance. *Journal of finance*, pages 57–82, 1997.

J. Conrad, R.F. Dittmar, and E. Ghysels. Ex ante skewness and expected stock returns. *Ann Arbor*, 1001:48109.

E.F. Fama and K.R. French. Size and book-to-market factors in earnings and returns. *Journal of Finance*, pages 131–155, 1995.

E.F. Fama and K.R. French. The equity risk premium. *Journal of Finance*, 57(2), 2002.

E.F. Fama and J.D. MacBeth. Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, pages 607–636, 1973.

M. Guidolin and A. Timmermann. Optimal portfolio choice under regime switching, skew and kurtosis preferences. *Review of Financial Studies*, 21:889–935, 2008.

P.R. Hansen and A. Lunde. Realized variance and market microstructure noise. *Journal of Business and Economic Statistics*, 24(2):127–161, 2006.

C.R. Harvey and A. Siddique. Conditional skewness in asset pricing tests. *The Journal of Finance*, 55(3):1263–1295, 2000.

C.R. Harvey, J. Liechty, M.W. Liechty, and P. Müller. Portfolio selection with higher moments. *Quantitative Finance*, 10(5):469–485, 2010.

J.C. Jackwerth and M. Rubinstein. Recovering probability distributions from option prices. *Journal of Finance*, pages 1611–1631, 1996.

E. Jondeau and M. Rockinger. The economic value of distributional timing. *Swiss Finance Institute Research Paper*, 2008.

X. Liu, M.B. Shackleton, S.J. Taylor, and X. Xu. Closed-form transformations from risk-neutral to real-world distributions. *Journal of Banking and Finance*, 31(5):1501–1520, 2007.

M.W. McCracken. Asymptotics for out-of-sample tests of causality. *manuscript, Louisiana State University*, 1999.

W.K. Newey and K.D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix, 1986.

M. Rubinstein. Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active cboe option classes from august 23, 1976 through august 31, 1978. *Journal of Finance*, pages 455–480, 1985.

M. Rubinstein. Implied binomial trees. *Journal of finance*, pages 771–818, 1994.

P. Santa-Clara and S. Yan. Crashes, volatility, and the equity premium: Lessons from s&p 500 options. *The Review of Economics and Statistics*, 92(2):435–451, 2010.

R.C. Scott and P.A. Horvath. On the direction of preference for moments of higher order than the variance. *The Journal of Finance*, 35(4):915–919, 1980.

S.J. Taylor. Asset price dynamics, volatility, and prediction. *Princeton University Press*, pages 423–466, 2007.

I. Welch and A. Goyal. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455, 2008.

Y. Xing, X. Zhang, and R. Zhao. What does the individual option volatility smirk tell us about future equity returns? *Journal of Financial and Quantitative Analysis*, 45(03):641–662, 2010.

7 Appendix

7.1 Methodology

Table 12: MSPE's of the daily AR models
This table reports the MSPE's of the daily AR models for different values of the estimation length and number of lags, K. The number of lags we use equal 1, 5, 20, 30, and 40 trading-days. The estimation lengths vary from 210 till 630 trading-days (10 till 30 months). The AR models are fitted on the log-second moment ($\ln(m_2)$), third moment (m_3), log-fourth moment ($\ln(m_4)$), log-volatility ($\ln$(vola)), skewness (skew), and log-kurtosis (ln(kurt)). The sample period extends from January 1990 to January 1996. For all models the predictions start at July 1992.

No. of lags	Estimation length (trading-days)	$\ln(m_2)$	m_3	$\ln(m_4)$	ln(vola)	skew	ln(kurt)
1	210	0.41	0.00	3.25	0.10	0.07	0.12
	252	0.42	0.00	3.15	0.10	0.07	0.12
	294	0.41	0.00	3.17	0.10	0.07	0.12
	336	0.42	0.00	3.24	0.10	0.07	0.12
	378	0.41	0.00	3.19	0.10	0.07	0.12
	420	0.41	0.00	3.18	0.10	0.07	0.12
	462	0.40	0.00	3.15	0.10	0.07	0.12
	504	0.41	0.00	3.17	0.10	0.07	0.12
	546	0.41	0.00	3.21	0.10	0.07	0.12
	588	0.41	0.00	3.18	0.10	0.07	0.12
	630	0.41	0.00	3.18	0.10	0.07	0.12
5	210	0.40	0.00	3.21	0.10	0.08	0.12
	252	0.41	0.00	3.15	0.10	0.08	0.12
	294	0.40	0.00	3.15	0.10	0.08	0.12
	336	0.41	0.00	3.22	0.10	0.08	0.12
	378	0.40	0.00	3.19	0.10	0.08	0.12
	420	0.40	0.00	3.16	0.10	0.08	0.12
	462	0.40	0.00	3.14	0.10	0.08	0.12
	504	0.40	0.00	3.14	0.10	0.08	0.12
	546	0.40	0.00	3.19	0.10	0.08	0.12
	588	0.40	0.00	3.17	0.10	0.08	0.12
	630	0.40	0.00	3.17	0.10	0.08	0.12
20	210	0.36	0.00	3.01	0.09	0.12	0.12
	252	0.37	0.00	3.07	0.09	0.11	0.12
	294	0.37	0.00	3.07	0.09	0.11	0.12
	336	0.37	0.00	3.04	0.09	0.11	0.12
	378	0.37	0.00	3.05	0.09	0.11	0.13
	420	0.37	0.00	3.05	0.09	0.11	0.13
	462	0.37	0.00	3.03	0.09	0.11	0.13
	504	0.37	0.00	3.06	0.09	0.11	0.13
	546	0.38	0.00	3.12	0.09	0.11	0.13
	588	0.37	0.00	3.11	0.09	0.11	0.13
	630	0.37	0.00	3.09	0.09	0.11	0.13
30	210	0.37	0.00	2.67	0.09	0.11	0.10
	252	0.37	0.00	2.74	0.09	0.10	0.10
	294	0.36	0.00	2.61	0.09	0.10	0.10
	336	0.35	0.00	2.60	0.09	0.10	0.10
	378	0.36	0.00	2.64	0.09	0.10	0.10
	420	0.35	0.00	2.62	0.09	0.10	0.10
	462	0.36	0.00	2.63	0.09	0.09	0.10
	504	0.35	0.00	2.60	0.09	0.09	0.10
	546	0.36	0.00	2.67	0.09	0.10	0.10
	588	0.36	0.00	2.68	0.09	0.09	0.10
	630	0.37	0.00	2.70	0.09	0.09	0.10
40	210	0.34	0.00	2.68	0.08	0.11	0.10
	252	0.35	0.00	2.76	0.09	0.10	0.10
	294	0.35	0.00	2.58	0.09	0.10	0.09
	336	0.34	0.00	2.55	0.08	0.10	0.10
	378	0.35	0.00	2.60	0.09	0.09	0.10
	420	0.34	0.00	2.58	0.09	0.10	0.10
	462	0.35	0.00	2.54	0.09	0.09	0.10
	504	0.34	0.00	2.55	0.08	0.09	0.10
	546	0.35	0.00	2.59	0.09	0.09	0.10
	588	0.35	0.00	2.59	0.09	0.09	0.09
	630	0.35	0.00	2.61	0.09	0.09	0.10

Table 13: MSPE's of the monthly AR models
This table reports the MSPE's of the monthly AR models for different values of the estimation length and number of lags, K. The number of lags we use equal 1, 5, 20, 30, and 40 days. The estimation lengths vary from 10 till 30 months. The AR models are fitted on the log-second moment ($\ln(m_2)$), third moment (m_3), log-fourth moment ($\ln(m_4)$), log-volatility ($\ln$(vola)), skewness (skew), and log-kurtosis (ln(kurt)). The sample period extends from January 1990 to January 1996. For all models the predictions start at July 1992.

No. of lags	Estimation length (months)	$\ln(m_2)$	m_3	$\ln(m_4)$	ln(vola)	skew	ln(kurt)
	10	0.29	0.00	2.27	0.07	0.10	0.11
	12	0.29	0.00	2.24	0.07	0.10	0.10
	14	0.28	0.00	2.25	0.07	0.09	0.10
	16	0.28	0.00	2.22	0.07	0.09	0.10
	18	0.28	0.00	2.21	0.07	0.09	0.10
1	20	0.28	0.00	2.19	0.07	0.09	0.10
	22	0.27	0.00	2.14	0.07	0.09	0.10
	24	0.27	0.00	2.14	0.07	0.09	0.10
	26	0.27	0.00	2.15	0.07	0.09	0.10
	28	0.27	0.00	2.13	0.07	0.09	0.10
	30	0.27	0.00	2.13	0.07	0.09	0.10
	10	0.33	0.00	2.63	0.08	0.12	0.11
	12	0.34	0.00	2.67	0.08	0.11	0.10
	14	0.32	0.00	2.28	0.08	0.10	0.10
	16	0.31	0.00	2.23	0.08	0.09	0.09
	18	0.30	0.00	2.26	0.07	0.09	0.10
2	20	0.29	0.00	2.19	0.07	0.09	0.09
	22	0.29	0.00	2.18	0.07	0.09	0.09
	24	0.28	0.00	2.13	0.07	0.09	0.09
	26	0.28	0.00	2.21	0.07	0.09	0.09
	28	0.28	0.00	2.15	0.07	0.09	0.09
	30	0.28	0.00	2.15	0.07	0.09	0.09
	10	0.44	0.00	3.80	0.11	0.15	0.14
	12	0.38	0.00	3.21	0.09	0.13	0.11
	14	0.37	0.00	2.60	0.09	0.12	0.10
	16	0.32	0.00	2.47	0.08	0.10	0.09
	18	0.33	0.00	2.43	0.08	0.10	0.09
3	20	0.30	0.00	2.38	0.08	0.10	0.09
	22	0.30	0.00	2.29	0.08	0.10	0.09
	24	0.29	0.00	2.17	0.07	0.09	0.08
	26	0.28	0.00	2.21	0.07	0.09	0.08
	28	0.29	0.00	2.20	0.07	0.09	0.08
	30	0.28	0.00	2.14	0.07	0.09	0.08
	10	0.88	0.00	8.33	0.22	0.26	0.24
	12	0.56	0.00	4.90	0.14	0.16	0.16
	14	0.49	0.00	4.31	0.12	0.14	0.11
	16	0.41	0.00	2.72	0.10	0.13	0.10
	18	0.36	0.00	2.61	0.09	0.11	0.09
4	20	0.33	0.00	2.59	0.08	0.10	0.10
	22	0.32	0.00	2.50	0.08	0.10	0.09
	24	0.31	0.00	2.25	0.08	0.10	0.09
	26	0.31	0.00	2.37	0.08	0.10	0.08
	28	0.30	0.00	2.30	0.07	0.10	0.08
	30	0.31	0.00	2.36	0.08	0.09	0.08
	10	185.58	0.01	82.59	52.96	43.57	0.04
	12	0.97	0.00	13.01	0.24	0.31	0.28
	14	0.58	0.00	6.72	0.14	0.20	0.18
	16	0.51	0.00	3.41	0.13	0.18	0.12
	18	0.39	0.00	3.13	0.10	0.13	0.11
5	20	0.40	0.00	2.95	0.10	0.11	0.10
	22	0.34	0.00	2.80	0.08	0.11	0.10
	24	0.33	0.00	2.74	0.08	0.11	0.09
	26	0.32	0.00	2.40	0.08	0.11	0.09
	28	0.31	0.00	2.34	0.08	0.10	0.09
	30	0.31	0.00	2.25	0.08	0.10	0.08

7.2 Results

7.2.1 Power Curves

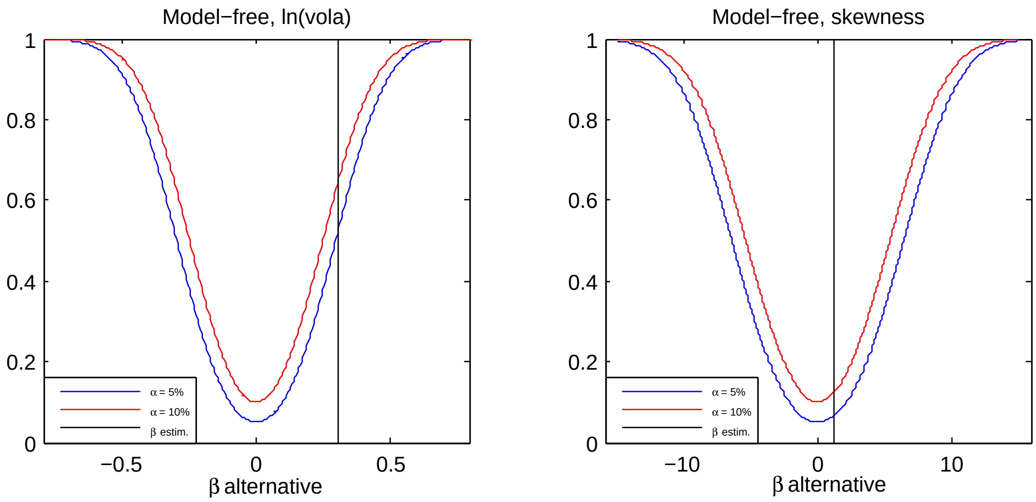

Figure 7: Power curve of the log-volatility and skewness coefficient for the model-free method.
This figure displays the power curve of the jointly regressed log-volatility (left) and skewness (right) coefficient for the model-free method, for a size, α, equal to 5 and 10 percent. The vertical line (in black) indicates the value of the estimated coefficient, β. It relies on the assumption of asymptotic normality.

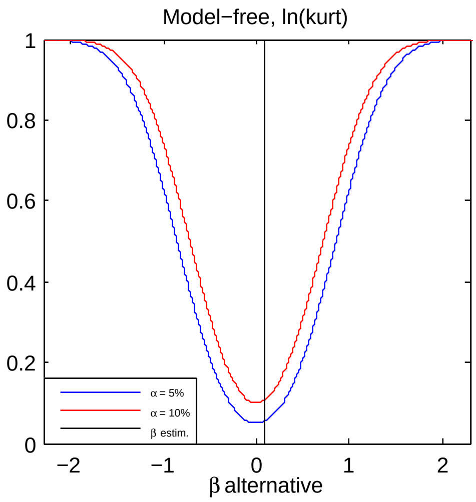

Figure 8: Power curve of the log-kurtosis coefficient for the model-free method.
This figure displays the power curve of the jointly regressed log-kurtosis coefficient for the model-free method, for a size, α, equal to 5 and 10 percent. The vertical line (in black) indicates the value of the estimated coefficient, β. It relies on the assumption of asymptotic normality.

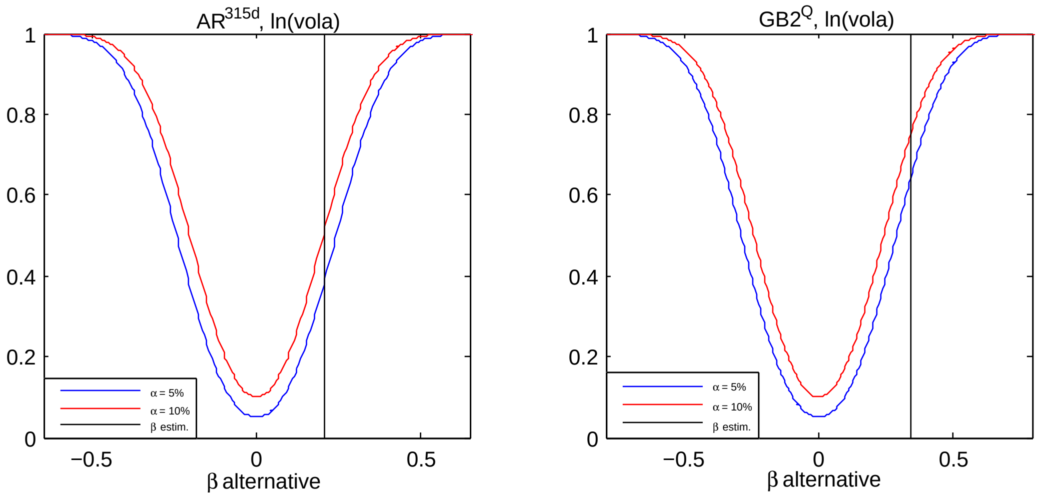

Figure 9: Power curve of the log-volatility coefficient for the AR^{315d} and $GB2^{Q}$ method.
This figure displays the power curve of the jointly regressed log-volatility coefficient for the AR^{315d} and $GB2^{Q}$ method, for a size, α, equal to 5 and 10 percent. The vertical lines (in black) indicates the value of the estimated coefficients, β. It relies on the assumption of asymptotic normality.

7.2.2 Results Alternative Realized Moments

Table 14: Summary statistics of the volatility, skewness, and kurtosis series, using the alternative realized moments
This table reports the summary statistical of the volatility, skewness, and kurtosis series as computed by the methods as stated in the column headings. The statistics are all based on monthly data, however, the volatility series is annualized. The methods considered are the realized method, daily autoregressive method, using a estimation length of 315 days (AR^{315d}), monthly autoregressive method, using a estimation length of 15 months (AR^{15m}), model-free option-implied risk-neutral method, and Generalized Beta distribution of the second kind option-implied risk-neutral ($GB2^{Q}$) and real-world ($GB2^{P}$) method. All AR(1) coefficients are statistically different from zero, at a (two-sided) one percent significance level. The sample period extends from January 1996 to March 2010.

	Volatility			Skewness			Kurtosis		
	Mean	Std. dev.	AR(1)	Mean	Std. dev.	AR(1)	Mean	Std. dev.	AR(1)
Realized*	0.18	0.12	0.81	0.03	0.22	0.25	7.64	3.72	0.79
AR^{315d*}	0.20	0.14	0.79	0.01	0.07	0.23	7.57	3.54	0.91
AR^{15m*}	0.20	0.14	0.84	0.00	0.11	0.08	7.80	3.87	0.81
Model-free	0.20	0.09	0.85	-1.38	0.37	0.58	6.35	1.87	0.55
$GB2^{Q}$	0.21	0.09	0.84	-1.18	0.35	0.42	5.93	1.05	0.52

Table 15: Correlation between moment series, using the alternative realized moments
This table reports the Pearson correlation coefficients between the (transformed) moment series as used in the regressions. The variables that we consider are the log-volatility (ln(vola)), skewness, and log-kurtosis (ln(kurt)). All coefficients are based on monthly data. The sample period extends from January 1996 to March 2010.

		skewness	ln(kurt)
Realized*	ln(vola)	-0.07	-0.32
	skewness		0.20
$AR^{315d}*$	ln(vola)	-0.04	-0.37
	skewness		0.19
$AR^{15m}*$	ln(vola)	-0.05	-0.34
	skewness		-0.02
Model-free	ln(vola)	0.13	-0.05
	skewness		-0.93
$GB2^{Q}$	ln(vola)	0.02	-0.36
	skewness		-0.89

Table 16: Static regression results, using the alternative realized moments

This table reports the regression coefficients of the log-volatility (ln(vola)), skewness, and log-kurtosis (ln(kurt)) variable as used in the static regression, for the different computational methods. Additionally, it reports the F-statistic on the significance of all three variables combined, and the R^2. The computational methods that we consider are the realized method, daily autoregressive method, using a estimation length of 315 days (AR^{315d}), monthly autoregressive method, using a estimation length of 15 months (AR^{15m}), model-free option-implied risk-neutral method, and option-implied risk-neutral Generalized Beta Distribution of the Second Kind method ($GB2^{Q}$). The results are generated using the regression model as specified in equation 34 till 36. The intercept and coefficients of the control variables are not reported. The Newey-West t-statistics are stated in the parenthesis. The one, five, and ten percent (two-sided) significance level is indicated with three, two, and one stars. The power results from the evaluation of the power curve, pertaining to the significance level as stated in the table, at the value of the estimated coefficient. It relies on an asymptotic normality. The sample period extends from January 1996 to March 2010, so the number of observations included in the regression equals 176.

	ln(vola)				skewness				ln(kurt)						
	coef.	t-stat	sign.	power (%)	coef.	t-stat	sign.	power (%)	coef.	t-stat	sign.	power (%)	F-stat.	sign.	R^2 (%)
Realized	0.25	(1.52)													3.24
					-0.60	(-0.31)									1.26
									-0.08	(-0.75)					1.42
	0.24	(1.32)			-0.25	(-0.12)			-0.01	(-0.06)			3.54	**	3.25
AR^{315d*}	0.17	(1.50)													2.13
					2.97	(0.66)									1.35
									-0.13	(-1.37)					1.82
	0.14	(1.14)			4.37	(0.95)			-0.10	(-0.93)			2.50	*	2.66
AR^{15m*}	0.24	(1.55)													3.31
					-1.09	(-0.43)									1.25
									-0.04	(-0.42)					1.27
	0.25	(1.41)			-0.65	(-0.24)			0.03	(0.23)			3.73	***	3.36
Model-free	0.31	(1.92)	*	61											3.07
					0.49	(0.48)									1.31
									-0.06	(-0.45)					1.30
	0.31	(2.05)	**	54	1.18	(0.36)			0.09	(0.20)			3.49	**	3.22
$GB2^{Q}$	0.26	(1.98)	**	51											2.90
					-0.04	(-0.04)									1.20
									-0.09	(-0.41)					1.27
	0.34	(2.31)	**	64	2.21	(0.48)			0.28	(0.32)			3.38	**	3.16

Table 17: Performance statistics, using the alternative realized moments
This Table displays the performance statistics of the five different trading strategies. The strategies are created by maximizing the expected utility each month. All figures are annualized ratios or annualized percentages, with the exception of the two conditional expectations, which are daily percentages. The results are based on a relative risk aversion parameter of 4. The sample period extends from January 1996 to March 2010.

	Realized*	AR^{315d*}	AR^{15m*}	Model-free	$GB2^{Q}$	S&P500
Mean weight	0.53	0.48	0.47	0.34	0.32	1.00
Std. dev. weight	0.40	0.39	0.39	0.23	0.23	0.00
Min weight	0.01	0.01	0.01	0.02	0.02	1.00
Max weight	1.80	1.70	1.65	1.08	1.10	1.00
Excess return	1.76	2.48	1.51	1.60	1.31	5.00
Std deviation	8.26	7.61	7.35	5.10	4.87	20.53
Downside risk	8.62	7.77	7.67	5.26	5.06	20.65
Correlation	0.69	0.68	0.65	0.76	0.74	1.00
Beta	0.28	0.25	0.23	0.19	0.18	1.00
Sharpe ratio	0.21	0.33	0.21	0.31	0.27	0.24
Sortino ratio	0.20	0.32	0.20	0.30	0.26	0.24
Treynor ratio (x10)	0.63	0.98	0.65	0.84	0.75	0.50
Alpha	0.38	1.21	0.35	0.65	0.43	0.00
M^2 measure	-0.77	1.59	-0.92	1.27	0.34	0.00
P($M^2 > 0$)	49.66	50.72	49.85	51.72	51.31	0.00
E[$M^2 \| M^2 > 0$]	0.57	0.60	0.61	0.47	0.50	0.00
E[$M^2 \| M^2 < 0$]	-0.57	-0.60	-0.62	-0.49	-0.52	0.00

Table 18: Performance of AR models, using the alternative realized moments
This table reports the performance statistics of the trading strategies based on different autoregressive (AR) models. These models are fitted on either daily or monthly data, and vary in the length of the estimation window. This length ranges from from 10 to 25 months (or 210 to 525 trading-days). The strategies are created by maximizing the expected utility each month. All figures are annualized ratios or annualized percentages. The results are based on a relative risk aversion parameter of 4. The sample period extends from January 1996 to March 2010.

	AR^{210d*}	AR^{315d*}	AR^{420d*}	AR^{525d*}	AR^{10m*}	AR^{15m*}	AR^{20m*}	AR^{25m*}
Mean weight	0.49	0.48	0.46	0.46	0.48	0.47	0.46	0.46
Std. dev. weight	0.39	0.39	0.38	0.38	0.39	0.39	0.38	0.38
Min weight	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Max weight	1.61	1.70	1.58	1.58	1.83	1.65	1.73	1.66
Excess return	2.11	2.48	2.48	2.54	1.58	1.51	1.37	1.41
Std deviation	7.85	7.61	7.44	7.40	7.55	7.35	7.24	7.19
Downside risk	8.10	7.77	7.57	7.51	7.91	7.67	7.57	7.50
Correlation	0.69	0.68	0.69	0.68	0.65	0.65	0.65	0.65
Beta	0.26	0.25	0.25	0.24	0.24	0.23	0.23	0.23
Sharpe ratio	0.27	0.33	0.33	0.34	0.21	0.21	0.19	0.20
Sortino ratio	0.26	0.32	0.33	0.34	0.20	0.20	0.18	0.19
Treynor ratio (x10)	0.80	0.98	1.00	1.04	0.66	0.65	0.60	0.62
Alpha	0.79	1.21	1.24	1.32	0.38	0.35	0.23	0.27
M^2 measure	0.40	1.59	1.77	1.98	-0.87	-0.92	-1.27	-1.13
P($M^2 > 0$)	50.47	50.72	50.77	50.39	50.58	49.85	50.12	49.72
E[$M^2 \| M^2 > 0$]	0.57	0.60	0.59	0.60	0.59	0.61	0.61	0.61
E[$M^2 \| M^2 < 0$]	-0.58	-0.60	-0.60	-0.59	-0.61	-0.62	-0.62	-0.61

7.3 Risk-Neutral Density from Option Prices

Introduction

When extracting the risk-neutral probability density function from option prices, often a parametric method is preferred. The reason is that the flexibility of the non-parametric methods often causes problems, such as badly behaved distributions and optimization problems can easily arise. the literature describes multiple parametric methods to extract the implied risk-neutral distribution from option prices. Most common parametric methods are lognormal mixtures, first postulated by Ritchey (1990)[10]; lognormal polynomials, developed by Madan and Milne (1994)[11], stochastic volatility; applied by, amongst others, Jondeau and Rockinger (2000)[12]; and the Generalized Beta Distribution of the Second Kind (GB2), as first proposed by Bookstaber and McDonald (1987 and 1991)[13]. Another approach is the method of parametric implied volatility functions (IVF), applied by Shimko (1993)[14], Campa et al. (1998)[15]. and Breeden and Lintenberger (1978)[16], amongst others.

Before selecting a method most suitable for this research, lets first go trough some useful criteria, as set out by Taylor (2007)[17]. The methods should preferably score high on all of the following criteria:

1. Estimated densities are never negative.
2. General levels of skewness and kurtosis are allowed.
3. The tails of the distribution are fat relative to a lognormal distribution.
4. There are analytic formulae to compute the densities and call prices.

[10] R.J. Ritchey. Call option valuation for discrete normal mixtures. Journal of Financial Research, 13(4):285 to 96, 1990.

[11] D.B. Madan and F. Milne. Contingent claims valued and hedged by pricing and investing in a basis. Mathematical Finance, 4(3):223 to 245, 1994.

[12] E. Jondeau and M. Rockinger. Reading the smile: the message conveyed by methods which infer risk neutral densities. Journal of International Money and Finance, 19(6):885 to 915, 2000.

[13] R.M. Bookstaber and J.B. McDonald. A general distribution for describing security price returns. Journal of Business, pages 401 to 424, 1987, and R.M. Bookstaber and J.B. McDonald. Option pricing for generalized distributions. Communications in Statistics, pages 4053 to 4068, 1991.

[14] D. Shimko. Bounds of probability. Risk, 6(4):33 to 37, 1993

[15] J.M. Campa, PH Chang, and R.L. Reider. Implied exchange rate distributions: evidence from OTC option markets. Journal of International Money and Finance, 17(1):117 to 160, 1998.

[16] D.T. Breeden and R.H. Litzenberger. Prices of state contingent claims implicit in option prices. Journal of business, pages 621 to 651, 1978.

[17] S.J. Taylor. Asset price dynamics, volatility, and prediction. Princeton University Press, pages 423 to 466, 2007.

5. Estimates are not sensitive to option price discreteness.

6. Solutions to the parameter estimation problem are easy to obtain.

For this research, criteria two is of utmost importance as those higher moment are also being tested. Criteria one and six are important as well, however; violation also largely dependents on the characteristics of the data at hand. Next, we go through the most common parametric methods.

Lognormal Mixtures

In stead of fitting a single lognormal distribution, as assumed by the Black-Scholes formula, this methods fits multiple lognormal distributions (mostly two or three). The method is intuitive when the price of the underlying, P_t, depends on a limited number future states (two or three). However, a large number of parameters need to be estimated ($3n + 2$, for n distributions). It violates criteria three, five, and six.

Lognormal Polynomial Density Functions

This method multiplies a lognormal density by a polynomial to transform it to a more representative function for the evolution of prices. The method has a strong theoretical foundations, however; negative densities can only be avoided by restricting the level of skewness and kurtosis. The method violates criteria one and two.

Stochastic Volatility

Numerical integration of a stochastic differential equation yields the desired risk-neutral density. The differential equation could for example be the volatility diffusion process of Heston (1993)[18]. It is numerically an advantageous approach as the parameters do not depend on the time-to-expiration, τ, meaning that more option prices can be used to estimate the model. A disadvantage is that it imposes a structure on the evolution of the prices of the underlying. Violation occurs on criteria two, four, and six.

GB2 distribution

[18] S.L. Heston. A closed form solution for options with stochastic volatility with applications to bond and currency options. Review of financial studies, 6(2):327, 1993.

The GB2 distribution is much more flexible and therefore better able to describe the evolution of asset prices than, for example, a lognormal distribution. In particularly, it copes well with describing the skewness and excess kurtosis. A disadvantage of the model is that the parameters have no economic interpretation. The method does not seem to violate any of the above criteria.

Parametric IVF
This method does not specify a parametric distribution, but a parametric shape to the implied volatility smile. A polynomial or spline is fitted trough the implied volatilities, by either minimizing the sum of squared errors between the implied and model volatilities, or implied and market call prices. It violates criteria one.

Since lognormal polynomials and stochastic volatility both violate criteria two, we exclude them as appropriate methods. The violation of the remaining criteria depend largely on the data used. Therefore, the remaining three methods are tested on the actual data first. In this experiment the same option prices are used as in the actual research, being the 30 day time-to-expiration options of the S&P 500 index.

Empirical Study

Parametric IVF
In deriving a closed form solution to the risk-neutral density, we follow Breeden and Lintzenberger (1978). First consider the formula to compute the price of an European call option,

$$c_t(X, \tau) = e^{-r_f \tau} \int_X^\infty (P_T - X) f(P_T \mid \theta) \mathrm{d}P_T \ , \tag{52}$$

with $f(X)$ the risk-neutral probability density function of the price of the underlying, P_T, at time, T, evaluated over different strike prices, X. The risk-free rate and time-to-expiration are represented by r_f and τ respectively.

When differentiating this call price twice with respect to the strike price, X, we obtain the following:

$$\frac{\delta^2 c_t(X, \tau)}{\delta X^2} = e^{r_f T} f_Q(X) \ . \tag{53}$$

Rearranging the formula, and we see that the risk-neutral density is a function of the second order derivative of the call price with respect to the strike price.

$$f_Q(X) = e^{r_f \tau} \frac{\delta^2 c_t(X, \tau)}{\delta X^2} \ . \tag{54}$$

By taking the Black-Scholes formula as our option pricing model, we first need to substitute the polynomial expression for the constant volatility, and take the second derivative. This polynomial links the volatility to the strike prices via its parameters. Those parameters are estimated by minimizing the sum of squared errors, as the following objective states:

$$\min_{\{\theta\}} \text{SSE} = \sum_{i=1}^{N} (\sigma_{implied}(X_i, \tau) - \sigma_{polynomial}(X_i, \tau \mid \theta))^2 \ , \tag{55}$$

with $\sigma_{implied}(X_i)$ the implied Black-Scholes volatility from observed option prices, for different strike prices, X_i. The implied volatility according to the polynomial structure is defined by $\sigma_{polynomial}(X_i \mid \theta)$, with θ being the parameters of the polynomial. In total we use N different strike prices in the optimization.

Figure 10 shows one of the estimated distributions. This distribution behaves properly, in the sense that it does not become negative and it intergrades to one. However, this is not always the case. Figure 11 shows a distribution which is not defined in the right tail and hence does not integrate to one.

When looking at the IVF of Figure 10, we see that the second order polynomial is apparently too rigid to capture the tiny upswing in the implied volatility for relatively large values of the strike price. When the polynomial is extrapolated, it eventually becomes negative. Since negative volatilities are not allowed, and thus eliminated, the distribution is also not defined on those regions. This problem occurs in 16 out of 176 times. Simply omitting those distributional statistics from the data creates biasses, since this problem does not occur randomly over the sample period, but during specific periods of high volatility.

A third order polynomial would be much more flexible. Hence, it is expected to capture the upswings at the end (and downswings at the beginning) of the implied volatilities. Figure 12 shows that this is indeed the case. Although the far right tail of the distribution is still not defined, the problem is much less severe than with the second order polynomial. However, a major drawback is that the vast majority of distributions become negative as a result of this modification. See for example the distribution in figure 13. Because of these problems, the polynomial IVF method is not considered an appropriate method for this research.

Lognormal Mixtures

Again consider the price of a European call option as defined in equation 52. When fitting multiple lognormal distributions, we need to substitute those for $f(P_T \mid \theta)$, with a mixture law. For two lognormal distributions this leads to

$$c_t(X,\tau) = e^{-r_f\tau}\int_X^\infty (P_T - X)[\,\alpha L(\mu_1,\sigma_1,P_T) + (1-\alpha)L(\mu_2,\sigma_2,P_T)\,]\mathrm{d}P_T\ , \tag{56}$$

with α the parameter determining the mixture law, and μ and σ the mean and standard deviation of the *normal* equivalent of the lognormal distributions[19]. Bahra (1997)[20] shows that this expression has the following closed form solution:

$$c_t(X,\tau) = e^{-r_f\tau}\{\alpha[e^{\mu_1+\frac{1}{2}\sigma_1^2}N(d_1) - XN(d_2)] \tag{58}$$

$$+ (1-\alpha)[e^{\mu_2+\frac{1}{2}\sigma_2^2}N(d_3) - XN(d_4)]\}\ , \tag{59}$$

with d_1 and d_2 defined as

$$d_1 = \frac{\ln(X) + \mu_1 + \sigma_1}{\sigma_1^2},\ d_2 = d_1 - \sigma_1\ , \tag{60}$$

$$d_3 = \frac{\ln(X) + \mu_2 + \sigma_2}{\sigma_2^2},\ d_4 = d_3 - \sigma_2\ . \tag{61}$$

Note that the formula does not contain the price of the underlying at time T. This is because the forward price, F_t, is equivalent to the mean of the implied risk-neutral density, and hence can be expressed in terms of its parameters.

The risk-neutral density can now be extracted from option prices by estimating the pa-

[19]Note that the mean and standard deviation of the normal distribution, μ and σ, are related to those of the lognormal distribution, m and v, in the following way:

$$m = \ln(\mu) - \frac{1}{2}\ln\left(1 + \frac{\sigma^2}{\mu^2}\right),\ \ v = \ln\left(1 + \frac{\sigma^2}{\mu^2}\right). \tag{57}$$

[20]B. Bahra. Implied risk-neutral probability density functions from option prices: theory and application. Bank of England Working Paper No 66, 1997

rameters of the lognormal distributions and mixture law. The objective in this optimization is to minimize the sum of squared differences between the model and market call prices, so

$$\min_{\{\theta\}} \text{SSE} = \sum_{i=1}^{N} (c_{market}(X_i, \tau) - c_{model}(X_i, \tau \mid \theta_i))^2 \ . \quad (62)$$

When fitting multiple lognormal distributions we need to make a tradeoff between flexibility of the distribution and amount of parameters we need two estimate. By fitting two lognormal distributions it already appears to be difficult to estimate all five parameters properly for all dates, as each optimization needs a specific set of restrictions. Since, we have to carry out 167 optimizations, this approach is not considered optimal.

For the purpose of illustration, figure 14 and 15 show a distribution that is well behaved and not well behaved.

GB2 distribution

In this case the GB2 distribution is substituted into the equation 52. Hence,

$$c_t(X, \tau) = e^{-r_f \tau} \int_X^{\infty} (P_T - X) f_{GB2}(P_T \mid \alpha, \beta, \rho, \nu) \mathrm{d}P_T \ . \quad (63)$$

When solving the internal, the following expression is obtained:

$$\begin{aligned} c_t(X, \tau) = F e^{-r_f tau} [1 - F_\beta(u(X, \alpha, \beta) \mid \rho + \alpha^{-1}, \nu - \alpha^{-1}] \\ - X e^{-r\tau} [1 - F_\beta(u(X, \alpha, \beta) \mid \rho, \nu)] \ . \end{aligned} \quad (64)$$

with F_t the futures price of the underlying, defined as $F_t = P_t e^{(r_f - q)\tau}$, with q the dividend yield. The function u is defined as $u(x, \alpha, \beta) = \frac{(x/\beta)^\alpha}{1+(x/\beta)^\alpha}$.

The risk-neutral density can now be extracted from option prices by estimating the free parameters in the cumulative beta distribution. The objective in this optimization is to minimize the sum of squared differences between the model and market call prices, as

$$\min_{\{\theta\}} \text{SSE} = \sum_{i=1}^{N} (c_{market}(X_i, \tau) - c_{model}(X_i, \tau \mid \theta_i))^2 \ . \quad (65)$$

No difficulties arise in estimating the parameters in the 167 different optimizations, and all distributions are well behaved. See Figure 16 for an example.

Conclusion

To conclude, there are multiple methods to estimated the risk-neutral density function from option prices, all with their own advantages and disadvantages. From the parametric methods, the polynomial IVF, lognormal mixtures, and the GB2 method are preferred as they allow for general values of the higher moments of the distribution. The polynomial IVF often results in undefined or negative distributions. When fitting lognormal mixtures it appears to be very hard, if not impossible, to specify general constraints in the optimizations. The GB2 suffers of no such disadvantages. Therefore, this method is used in the research.

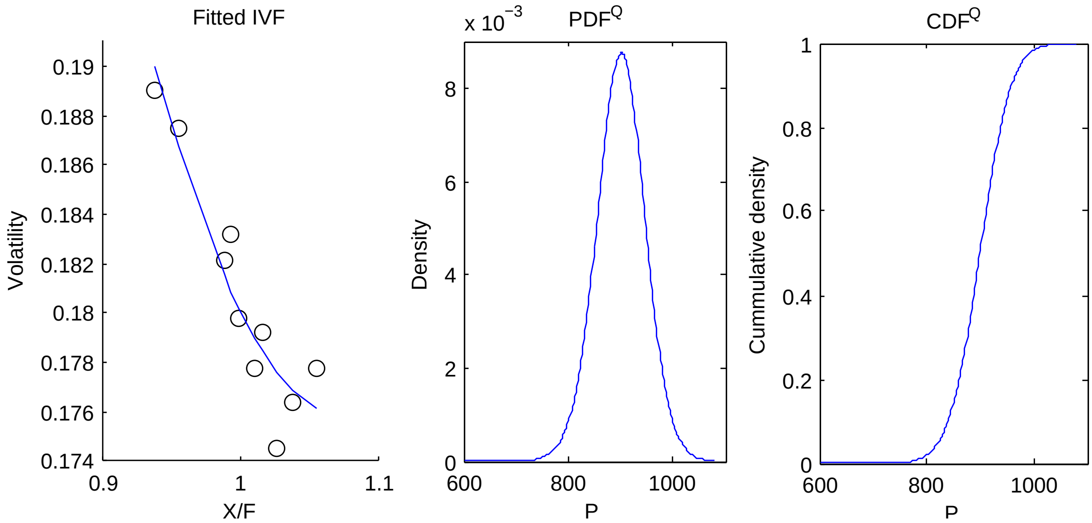

Figure 10: IVF approach, second order polynomial
This figure displays both risk-neutral probability density function (PDF^Q) and the cumulative probability density function (CDF^Q) of the price of the S&P 500 index, using the implied volatility function (IVF) approach. The function we use is a second order polynomial. In the optimization we use option data from 19 June 1997.

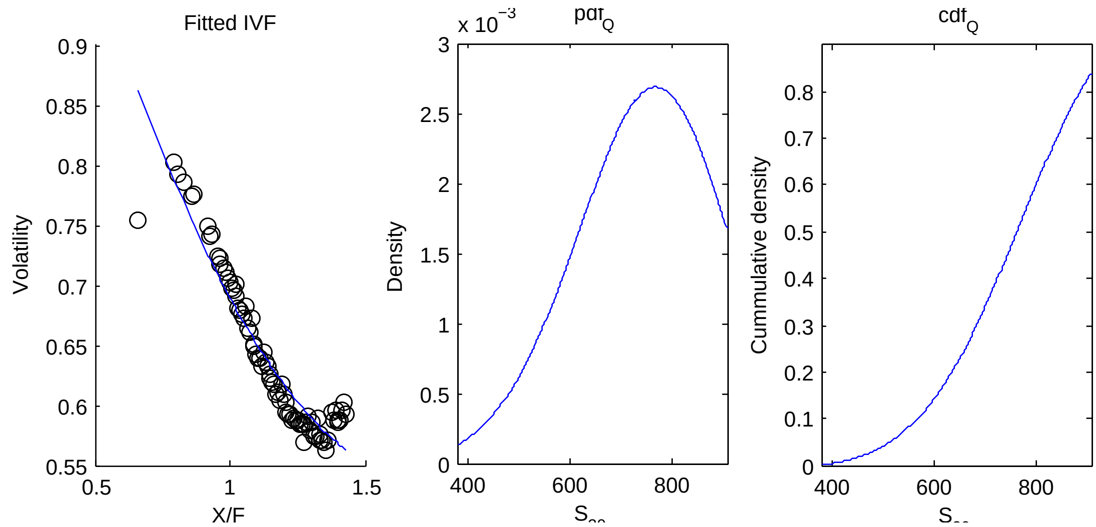

Figure 11: IVF approach, second order polynomial
This figure displays both risk-neutral probability density function (PDF^Q) and the cumulative probability density function (CDF^Q) of the price of the S&P 500 index, using the implied volatility function (IVF) approach. The function we use is a second order polynomial. In the optimization we use option data from 17 August 2000.

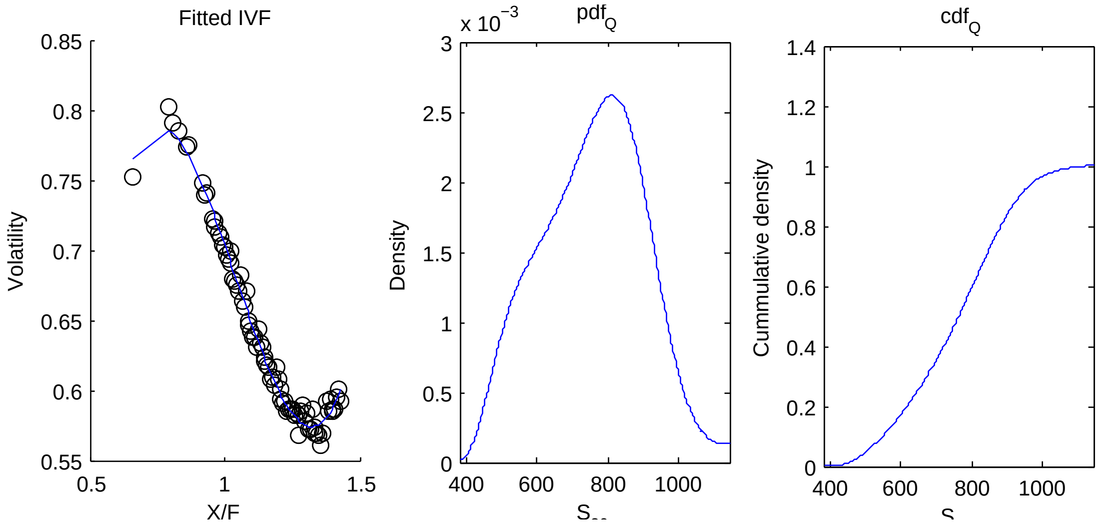

Figure 12: IVF approach, second order polynomial
This figure displays both risk-neutral probability density function (PDF^Q) and the cumulative probability density function (CDF^Q) of the price of the S&P 500 index, using the implied volatility function (IVF) approach. The function we use is a third order polynomial. The distribution is fitted at 19 June 1997.

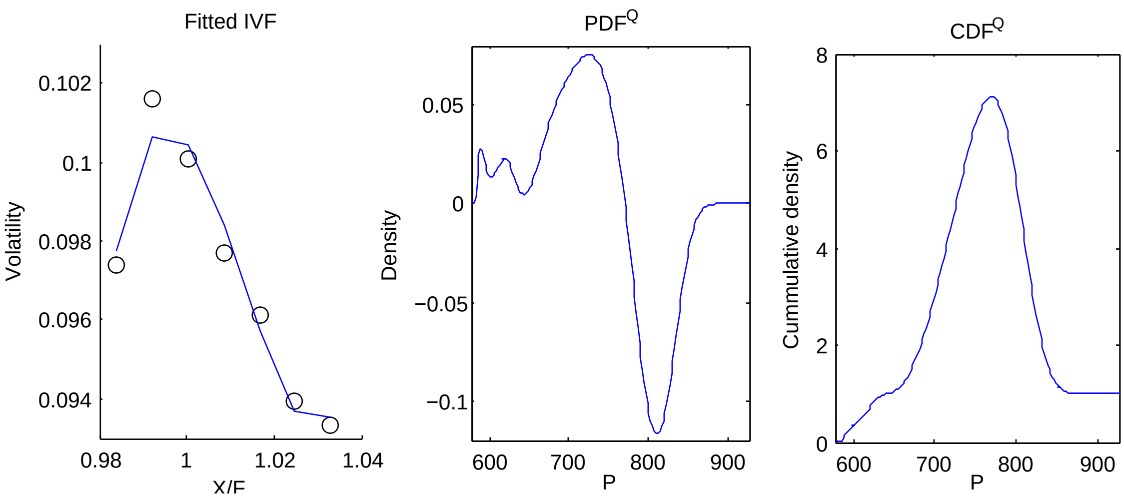

Figure 13: IVF approach, second order polynomial
This figure displays both risk-neutral probability density function (PDF^Q) and the cumulative probability density function (CDF^Q) of the price of the S&P 500 index, using the implied volatility function (IVF) approach. The function we use is a third order polynomial. The distribution is fitted at 18 January 1996.

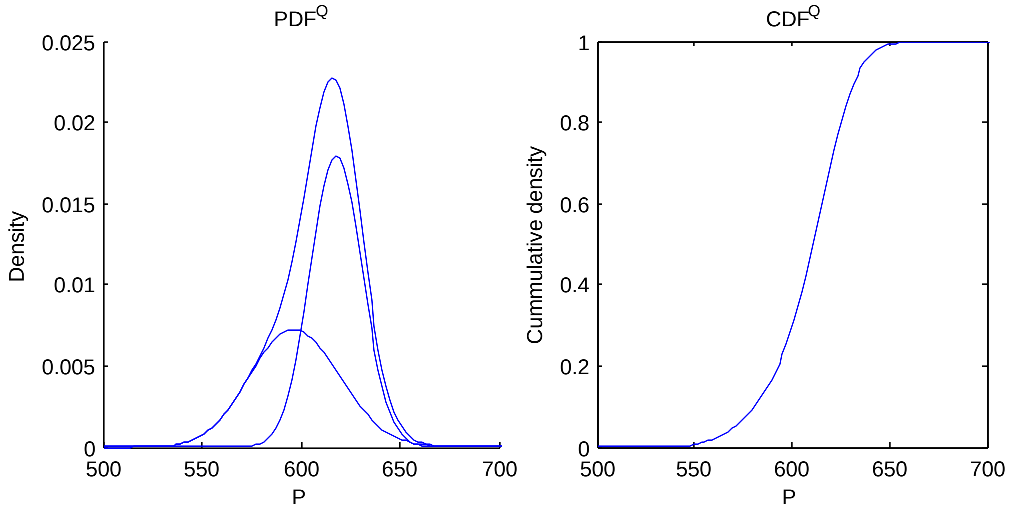

Figure 14: Lognormal mixtures
This figure displays both risk-neutral probability density function (PDF^Q) and the cumulative probability density function (CDF^Q) of the price of the S&P 500 index, by fitting two lognormal distributions. The distribution is fitted at 18 January 1996.

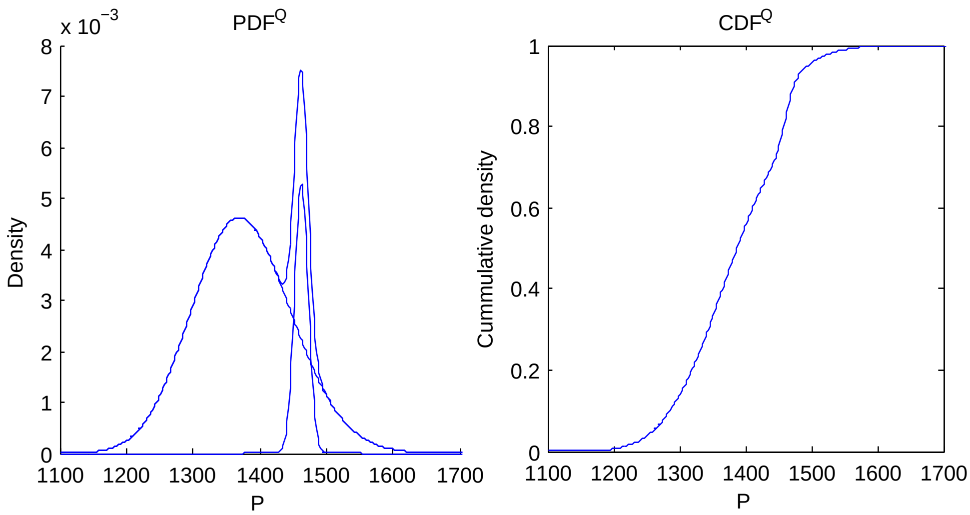

Figure 15: Lognormal mixtures
This figure displays both risk-neutral probability density function (PDF^Q) and the cumulative probability density function (CDF^Q) of the price of the S&P 500 index, by fitting two lognormal distributions. The distribution is fitted at 17 February 2000.

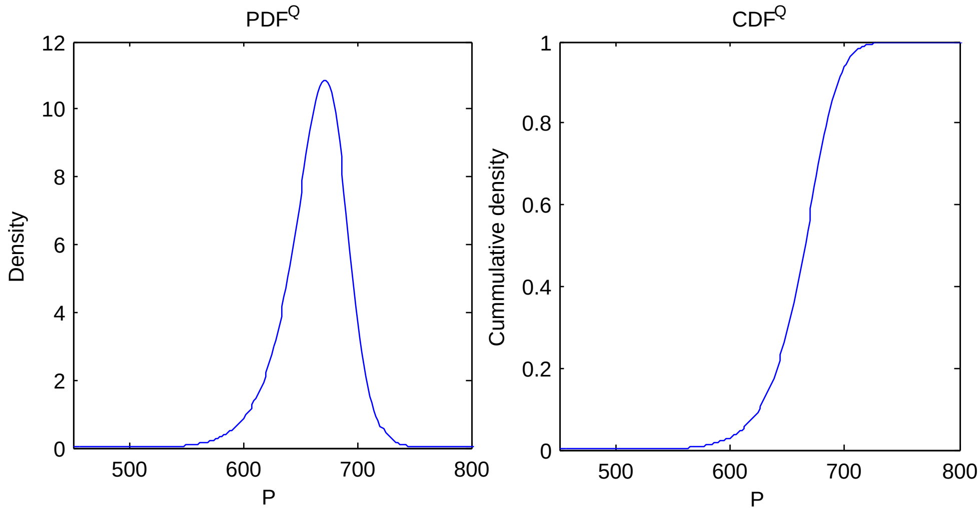

Figure 16: GB2 distribution
This figure displays both risk-neutral probability density function (PDF^Q) and the cumulative probability density function (CDF^Q) of the price of the S&P 500 index, by fitting a Generalized Beta distribution of the second kind (GB2). The distribution is fitted at 19 June 1997.