QUANTITATIVE FINANCE
RESEARCH CENTRE

UNIVERSITY OF
TECHNOLOGY SYDNEY

QUANTITATIVE FINANCE RESEARCH CENTRE

Research Paper 363 September 2015

Recovering the Real-World Density and Liquidity Premia From Option Data

Mathias Barkhagen, Jörgen Blomvall and Eckhard Platen

ISSN 1441-8010 www.qfrc.uts.edu.au

RECOVERING THE REAL-WORLD DENSITY AND LIQUIDITY PREMIA FROM OPTION DATA

MATHIAS BARKHAGEN, JÖRGEN BLOMVALL, AND ECKHARD PLATEN

Abstract. In this paper we develop a methodology for simultaneous recovery of the real-world probability density and liquidity premia from observed S&P 500 index option prices. Assuming the existence of a numéraire portfolio for the US equity market, fair prices of derivatives under the benchmark approach can be obtained directly under the real-world measure. Under this modeling framework there exists a direct link between observed call option prices on the index and the real-world density for the underlying index. We use a novel method for estimation of option implied volatility surfaces of high quality, which enables the subsequent analysis. We show that the real-world density that we recover is consistent with the observed realized dynamics of the underlying index. This admits the identification of liquidity premia embedded in option price data. We identify and estimate two separate liquidity premia embedded in S&P 500 index options that are consistent with previous findings in the literature.

1. Introduction

We recover the probabilistic description of the underlying index from option data. For liquid options at short maturities our method recovers the real-world probability density of the index. We also recover, for far out-of-the-money and longer term options, information which leads to a liquidity premium with characteristics that can probably be explained by observed trading behaviour. The model that we use to model the underlying index is a generalized version of the Minimal Market Model (MMM), which is described in Platen and Rendek (2012a) and Rendek (2013). The model has interesting properties that appear to support empirically observed phenomena that go beyond the classical no-arbitrage theory. In our setting we have that when the index has a major drawdown, the volatility shows extreme spikes which translates into explosion type features. In Rendek (2013) it is through extensive statistical tests shown that the generalized MMM provides a reasonably accurate description of the statistical characteristics of broad market equity indices. In this paper we also show that the real-world density estimated from option implied MMM parameters, is able to predict the statistical characteristics of the underlying index. This means that we are not only able to recover forward looking information, but also connect to the historically observed evolving index and its estimated characteristics. This in turn lets us discover the liquidity premia embedded in index option prices.

The recovery of the real-world probability density from liquid option prices relies on high quality estimation of the option implied risk-neutral density. Being able to estimate option implied risk-neutral densities, which are both realistic and consistent with observed option prices, is a crucial component of the estimation methodology. The risk-neutral densities are estimated with a recently developed non-parametric method (Barkhagen

JEL Classification. C61; C63; G12; G13.

Key words and phrases. Real-world density; Liquidity premia; Local volatility model; Non-parametric estimation; Simulated Maximum Likelihood.

and Blomvall, 2015), which has proven to produce densities that are realistic, consistent with market prices and stable over time. The method is, thus, able to adequately weigh these conflicting objectives which is very difficult to achieve. Using the results derived in Heath and Platen (2006) we can then directly recover the real-world probability density from the estimated risk-neutral densities. From the recovered real-world densities, the generalized MMM is then shown to be able to separate the liquidity premia from the densities, producing a more realistic estimate of the option implied real-world density of the underlying index level. The estimated liquidity premia are consistent with previous findings in the literature (Bollen and Whaley, 2004; Gârleanu *et al.*, 2009).

The novelty of the paper is that we are able to recover the real-world dynamics of the underlying index together with extra characteristics of the options market. In particular, this means that less liquid option prices are more expensive than theoretical prices, and this effect can be explained by the liquidity premia. This means that the recovery theorem given in Ross (2015) only applies to liquid options. Thus our modeling framework explains potentially more than the related papers on recovery in the literature; see Carr and Yu (2012) and Waldén (2013). It allows us to recover the real-world probability density and liquidity premia in the case when the underlying is a well-diversified index.

The rest of the paper is organized as follows. In Section 2 we present the main results of the paper. The modeling framework of the paper and the MMM are then presented in Section 3. In Section 4 we demonstrate the different estimation techniques that we use and the paper is finally concluded in Section 5.

2. Main results

We will here present the main results of the paper. In the first part of this section we will illustrate empirically, through an example using observed S&P 500 index options for a particular date, that our method for recovery of the risk-neutral density (RND) from market prices produces estimates of high quality. Using the estimated RND we will then illustrate how this density can be transformed to the real-world density of the underlying index using the method described in Heath and Platen (2006). From the recovered real-world density the parameters of the generalized MMM, described in Platen and Rendek (2012a), are then estimated with a simulation based Maximum Likelihood method. In the second part of this section we will present the results of a time series study of estimated MMM parameters implied by S&P 500 index option prices, and present results that identify liquidity premia embedded in S&P 500 index options.

2.1. **Empirical estimation under the risk-neutral measure.**

In order to be able to recover the real-world probability density function (PDF) from option prices, it is important to use an estimation method that is able to estimate realistic option implied volatility surfaces that are consistent with observed market prices. The estimation method, which we describe later on in detail, should, thus, be able to separate out the noise in market data in order to produce realistic surfaces. In Section 4.1 we present the non-parametric method that we use for the RND recovery from option prices. In order to illustrate the performance of the estimation method that we use, we will in this section present estimation results for a given set of observed option prices. We estimate the option implied RND from observed S&P 500 options end of day bid and ask quotes as of 3 October 2014. The description of the data set that we used for the estimation is included in the appendix. The recovery of the option implied real-world PDF from the estimated RND relies on the method described in Section 4.2. Since there exists a liquidity

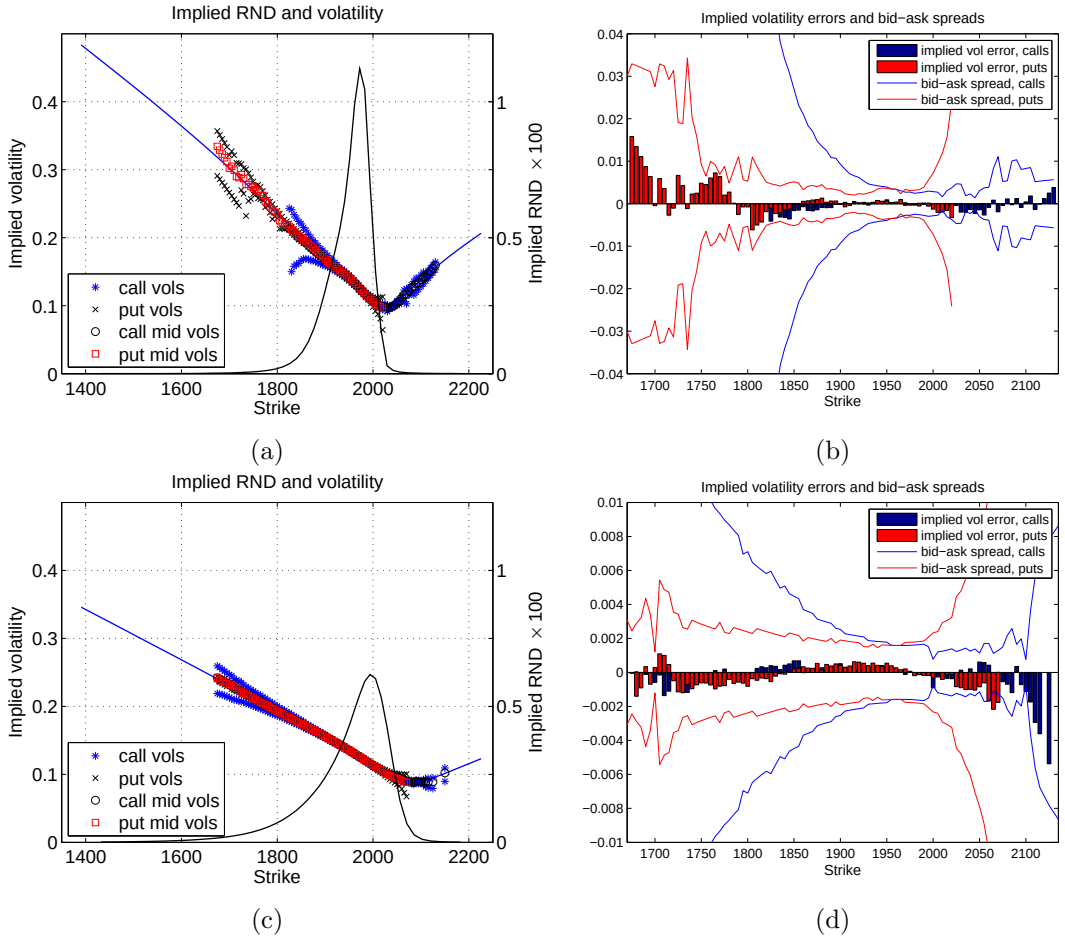


FIGURE 1. Figures (a) and (c) show the implied RND and volatility curves for the two shortest maturities (13 and 48 days to maturity) estimated from the S&P 500 index options data set as of 3 October 2014. The plots also include implied volatilities for the observed bid, ask and mid prices. Figures (b) and (d) show the implied volatility errors, measured as the difference between the estimated and mid-market implied volatilities, for the two shortest maturities. These plots also include the bid-ask spread for call (blue) and put (red) options for the different strikes in the data set.

premium for options with strikes that are much lower than the current index level (see e.g. Gârleanu *et al.*, 2009), and for long maturities (see e.g. Platen, 2009), we have to be careful when backing out the real-world PDF from observed option prices. Thus, we only use a subset, $\mathcal{L}$, of the observed options quotes, corresponding to liquid options close to at-the-money and close to maturity when backing out the RND. We have defined this subset as consisting of options with moneyness

$$m = \frac{K}{S_0} \in [0.85, 1.15], \tag{1}$$

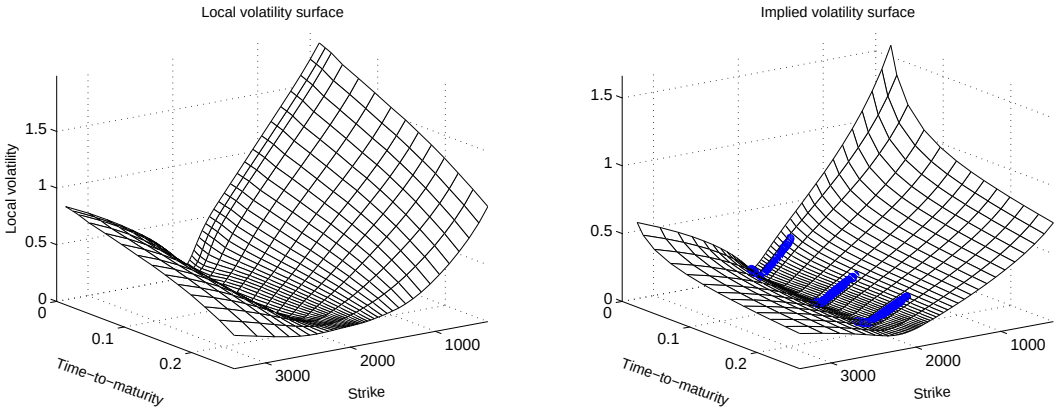
where S_0 is the current index level, K the strike price and the time-to-maturity $\tau \in [0, 91]$ in calendar days.

Figures 1a and 1c show the estimated RND and implied volatility for the two nearest maturities as of 3 October 2014. The spot index level is 1967.90 and the time-to-maturities

for these implied curves are 13 and 48 calendar days, respectively. As we can see from the plots, our estimation method produces smooth RNDs, as well as smooth and realistic implied volatility curves, that are consistent with observed option prices. We will use the estimated RNDs in order to recover the real-world PDFs that are implied by observed option prices. It is thus important to be able to weigh the conflicting objectives of producing local volatility and implied volatility surfaces that are smooth and realistic, and also consistent with observed option prices. If the option implied RND were not smooth, then this would imply that the estimation in part explains the noise present in the option data. If the level of market consistency were too low, then the option implied RND would not represent an accurate description of the information embedded in option prices. In order to investigate the pricing errors in terms of implied volatilities for these two maturities we have included their plots in Figures 1b and 1d. In these plots, the blue and red bars represent the implied volatility errors for calls and puts, respectively. These figures also include the bid-ask spreads in terms of implied volatility for the different strikes with observed option prices for these two maturities. As we can see from the plots, the implied volatility errors are generally much smaller than the corresponding spreads, and there is no apparent significant systematic pricing error visible. The mean absolute implied volatility errors (MAE) for call options for the data set is 0.06%, whereas the same measure for the put options for this data set is 0.11%.

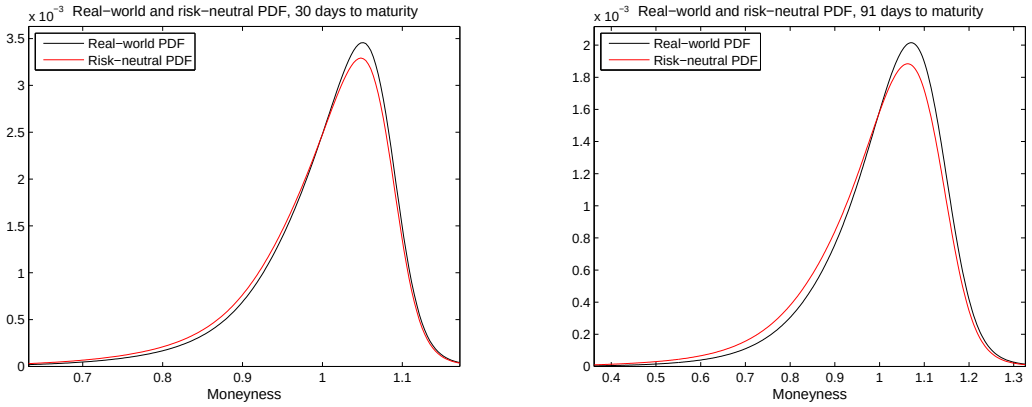
Figure 2a shows the estimated local volatility surface implied by the options in the set $\mathcal{L}$ as of 3 October 2014, and Figure 2b shows the corresponding implied volatility surface, as will be explained later. Even though there is a large amount of noise contained in the observed option quotes, our estimation method is able to produce a smooth and plausible local volatility surface. Constructing a smooth and plausible local volatility surface that is consistent with observed market prices is a difficult problem when observed prices are noisy. The typical shape of estimated local volatility surfaces from the literature includes spikes and is highly irregular and non-smooth (see e.g. Andersen and Brotherton-Ratcliffe, 1997; Coleman *et al.*, 1999). Being able to produce option implied surfaces of high quality is vital for the recovery of a plausible real-world PDF from option prices.

2.1.1. *Recovery of the real-world density.* From the estimated RNDs for different maturities, shown in the previous section, we can recover the real-world PDF with the method developed by Heath and Platen (2006). Under the assumption that a broad equity market index approximates the growth optimal portfolio, or numéraire portfolio (see Long, 1990), Heath and Platen (2006) show that there is a direct link between European call option prices written on the index and the real-world PDF. Note that there is no need to assume that the underlying index follows a local volatility function model, since we recover the unconditional, or marginal, densities from European call option prices. Furthermore, a result by Gyöngy (1986) ensures that there exists a projection of a multicomponent index model into a one-dimensional, scalar model. The real-world pricing of European call options to be presented in Section 3.2 will not rely on assuming a local volatility function model for the underlying index. Given the well-known link between call option prices and the RND (see Breeden and Litzenberger, 1978), the connection between the real-world PDF and the RND follows immediately when the index is assumed to be a good approximation of the numéraire portfolio. As in Heath and Platen (2006), we assume that the S&P 500 index is a good approximation of the numéraire portfolio for the US equity market, which allows us to recover the real-world PDF from observed prices of S&P 500 index options. These results will be derived in Section 4.2.



(a) (b)

FIGURE 2. Figure (a) shows the local volatility surface estimated from the S&P 500 index options end of day data as of 3 October 2014. Figure (b) shows the corresponding implied volatility surface, and the circles in the plot represent the mid-market implied volatilities for the options with observed bid and ask quotes.



(a) (b)

FIGURE 3. Figure (a) shows the option implied risk-neutral density (red) with 30 days to maturity, compared with the corresponding real-world density (black) estimated from the S&P 500 index options end of day data as of 5 August 2011. Figure (b) shows the risk-neutral and real-world densities with 91 days to maturity for the same data set.

In Figures 3a and 3b we illustrate the recovery of the real-world PDF from the S&P 500 index option data set described in Section 2.1, corresponding to 30 and 91 days to maturity. As is apparent from the figures, the left tail for the recovered real-world PDF (black curve) exhibits a lesser degree of *fat-tailedness*, or leptokurtosis, than the corresponding RND (red curve). The absolute value of the negative skewness is also lower for the real-world PDF than for the RND.

2.2. **Option implied parameters under the real-world measure.** The generalized Minimal Market Model (MMM) will be used in this paper as a tool for identification and estimation of liquidity premia present in the S&P 500 index option market. The

generalized MMM has in Platen and Rendek (2012b) and Rendek (2013) been shown to be able to rather accurately capture the empirically observed characteristics of a broad market index. By estimating the parameters of the generalized MMM from option data, the model will then explain the information embedded in the option data which is consistent with observed behaviour of the underlying market. The model will, however, not be able to explain the information embedded in option prices which exists due to supply and demand effects in the option market. Thus, by comparing the difference between observed market prices and prices implied by the fitted MMM parameters we can get an estimate of the liquidity premia embedded in S&P 500 index options.

The generalized MMM, developed in Platen and Rendek (2012b), models the savings account discounted total return index $\bar{S}_t$ at time $t \geq 0$ given by

$$\bar{S}_t = \frac{\tilde{S}_t}{B_t}, \tag{2}$$

where $\tilde{S}_t$ denotes the S&P 500 total return index and B_t the US savings account. The long term average excess growth of the discounted index $\bar{S}_t$ is modeled by an exponential function A_{τ_t} expressed in market activity time τ_t, which will be specified below, according to

$$A_{\tau_t} = A\,\mathrm{e}^{a\tau_t}, \tag{3}$$

for $t \geq 0$, $a \geq 1$, $A > 0$. The discounted total return index is then in the generalized MMM defined by

$$\bar{S}_t = A_{\tau_t} Y_{\tau_t}, \tag{4}$$

where Y_{τ_t} is the normalized index, which is evolving in the market activity time τ_t.

Motivated by economic arguments given in Platen and Rendek (2012a) the normalized index level is well modeled by a square root process $Y = \{Y_{\tau_t}, \tau_t \geq 0\}$ of dimension four in τ-time according to

$$dY_{\tau_t} = (1 - Y_{\tau_t})\, d\tau_t + \sqrt{Y_{\tau_t}} d\tilde{W}_{\tau_t}, \tag{5}$$

for $\tau_t \geq 0$, $Y_0 \geq 0$ and where $\tilde{W}_{\tau_t}$ denotes a Wiener process in τ-time. The only parameter that enters equation (5) is the initial value Y_0. The market activity time τ_t is here modeled via the ordinary differential equation

$$\frac{d\tau_t}{dt} = M_t, \tag{6}$$

where M_t is the adapted market activity, which will be (endogenously) specified below. The market activity is introduced in order to reflect human behaviour in economic activity and trading, which exaggerates the "normal" movements of volatility in market time in reaction to moves of the normalized index. It is well documented in the literature that trading activity increases when the market drops, whereas trading activity decreases when the index level increases (see e.g. Ané and Geman, 2000). This effect can be modeled by speeding up or slowing down, with respect to calendar time, the time scale under which the normalized index Y_{τ_t} evolves. This change in the market activity M_t is conveniently modeled by introducing the market activity time τ_t which by equation (6) is defined as the integrated market activity.

From equation (5) we see that the variance of the square root process Y is given by $1/Y_{\tau_t}$. As argued by Platen and Rendek (2012a), the market activity is therefore, similar

to the variance of Y, modeled as the inverse of a square root process. Thus, the process

$$X_t = \frac{1}{M_t}, \tag{7}$$

which is the inverse of the market activity process M_t, is modeled by a square root process of dimension four in t-time according to the stochastic differential equation (SDE)

$$dX_t = (\gamma - \epsilon X_t)\, dt + \sqrt{\gamma X_t} dW_t = \epsilon \left(\frac{\gamma}{\epsilon} - X_t \right) dt + \sqrt{\gamma X_t} dW_t, \tag{8}$$

where $t \geq 0$, $M_0 > 0$, $\gamma > 0$, $\epsilon > 0$ and where W_t denotes a Wiener process in t-time. From equation (8) we see that the speed of mean reversion for the inverse market activity is given by ϵ, and that the mean reversion level is given by $\frac{\gamma}{\epsilon}$. The Wiener process $\tilde{W}_{\tau_t}$, which models the non-diversifiable uncertainty driving Y_{τ_t} in market activity time, is linked to the Wiener process W_t in t-time according to the SDE

$$d\tilde{W}_{\tau_t} = \sqrt{\frac{d\tau_t}{dt}} dW_t = \sqrt{M_t} dW_t, \tag{9}$$

for $t \geq 0$ and $W_0 = 0$. From (2)-(6) and (9) we conclude that the instantaneous MMM volatility for the index at time t is given by

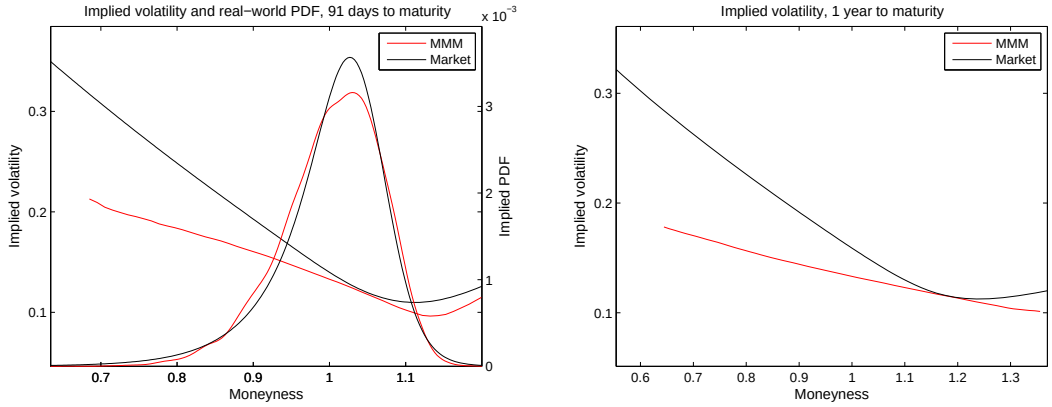
$$\sigma_t = \sqrt{\frac{M_t}{Y_{\tau_t}}} \tag{10}$$

The proposed index model is a Markovian two-component model, driven by only one Wiener process W_t. The two components are the normalized index Y_{τ_t} and the inverse market activity X_t. The three initial and three structural parameters of the model are: the parameter $A \geq 0$ for fitting the initial value of the average exponential growth part; the initial value $M_0 > 0$ of the market activity; the initial value $Y_0 > 0$ of the square root process; the long term average excess growth rate $a \geq 1$ of the discounted index; the speed of mean-reversion parameter $\epsilon > 0$ for the inverse market activity and the scaling parameter $\gamma > 0$ in the diffusion coefficient of the SDE (8) for the inverse of the market activity.

When estimating the option implied MMM parameters we use historical data of the discounted total return index in order to estimate A, a and Y_0, using the method described in Platen and Rendek (2012a). The reason for estimating these parameters from historical data is that the estimates allow us to steer the optimization towards plausible values. The remaining three parameters, ϵ, γ and M_0, which are implied by the recovered real-world PDF, are then estimated using Simulated Maximum Likelihood (SML) to be described in Section 4.3. In order to extract option implied information for the MMM parameters for both short term and longer term maturities, we will use the recovered option implied PDF for 5, 30 and 91 days as input to the SML algorithm. The resulting optimization problem that we solve is non-convex with multiple local optima. We handle the multiple local optima in the time series study that we present in Section 2.3 by generating multiple starting values for the model parameters for the first estimation date, and choose the parameter set corresponding to the solution with the largest objective value as the solution for this date. An example of estimated parameter values is given in Table 1, which contains the parameter values for ϵ, γ and M_0 implied by S&P 500 index options data as of 3 October 2014.

TABLE 1. MMM parameter values estimated from S&P 500 option data as of 3 October 2014. The parameter values for A, a and Y_0 are estimated from historical data of the discounted total return index from January 1988 to October 2014.

Parameter	ϵ	γ	M_0	A	a	Y_0
Estimate	62.1	6213	0.0344	1732	1.44	1.14



(a) (b)

FIGURE 4. Figure (a) compares the market (black) and MMM (red) real-world PDF and implied volatility with 91 days to maturity estimated from S&P 500 option prices on 3 October 2014. Figure (b) compares the corresponding market and MMM implied volatilities with 1 year to maturity.

These parameter values were then used to simulate values for the total return index level in order to price derivatives on the S&P 500 index under the benchmark approach, as will be described in Section 3.2. Figures 4a and 4b show the market implied volatilities compared to the MMM implied volatilities for maturity dates which lie 91 days and one year into the future, respectively. Figure 4a also includes the real-world density implied by market prices and the corresponding MMM density. From these plots we observe that option prices implied by the fitted MMM parameters are lower than the observed market prices for strikes that are far away from the current index level. This price discrepancy in the tails can be explained by a liquidity premium that stems from excess demand for especially out-of-the-money put options (see e.g. Gârleanu *et al.*, 2009). Institutions use out-of-the-money index puts in order to hedge their equity portfolios and this creates excess demand, which drives the price up for these options. In Platen (2009), it is shown that the price of a long term European put option under the risk-neutral measure is higher than the fair put option price obtained under the benchmark approach. Inspecting Figure 4b, which shows the implied volatility curves for options with one year to maturity, we see that the market prices are consistently higher than the prices implied by the MMM parameters. This is an indication that the option implied MMM parameters can be used in order to identify the term premium included in the market prices of options with longer maturities.

2.3. **Time series of option implied parameters.** In this section we present the time series of the option implied MMM parameters, estimated from the PDFs that are implied by historical closing prices of S&P 500 index options over the period 3 January 2011 to 7 October 2014. This means that we are estimating the implied parameters over 947 historical trading days. The parameters A, a and Y_0 have for each date been estimated from time series data of the discounted total return index, and hence we are solving the estimation problem for the following three model parameters: $\{\epsilon, \gamma, M_0\}$. We start, for the first date in the time series, estimating the option implied MMM parameters for 20 randomly chosen starting values for the model parameters. We then choose the set of parameters corresponding to the solution with the largest objective value as the solution for the first date. For all the following dates in the time series we then use the parameters from the previous day as starting values in the estimation.

Before we present the estimation results we will first motivate how we redefine the parameter set in order to get a more stable optimization problem when backing out the option implied MMM parameters. In Platen and Rendek (2012a) it has been shown that the long term average value of the squared instantaneous volatility is approximately given by

$$\sigma_\infty^2 \approx 4\frac{\epsilon}{\gamma}, \tag{11}$$

which signals ϵ/γ as being a crucial parameter. Indeed, we realized when estimating the option implied MMM parameters with SML, the optimization attempts to match the volatility for the shortest maturity by adjusting the parameter M_0, and simultaneously aims at matching the 91 day volatility by changing the parameter ϵ and the ratio ϵ/γ. For this reason we redefined the parameter set that we estimate with the SML algorithm from $\{\epsilon, \gamma, M_0\}$ to $\{\epsilon, \epsilon/\gamma, M_0\}$. This leads to an optimization problem with better properties than the original optimization problem.

When running a preliminary time series estimation of the MMM parameters, using the original parameter set, we observed that the optimization terminated at estimates for γ that either lay in the interval $(50, 10^3)$, or in an interval with much higher values, $(10^3, 10^4)$. From Table 1 the latter result is most likely the on average realistic one. Nevertheless, when generating the 20 randomly chosen starting values, we generated 10 random values in each of the two intervals. These starting values were then expressed as starting values for $\{\epsilon, \epsilon/\gamma, M_0\}$, and we then chose the parameters corresponding to the solution with the largest objective value as the solution for the first date. Since the optimization problem is non-convex with multiple local optima, and we use the solution from the previous day as starting value for the estimation, there is a risk that the estimation "gets stuck" at a local optimum that is close to the solution from the previous period. From the time series plot for the estimate of γ in Figure 5d, however, we see that the solution can jump to values that lie far from the solution of the previous day.

From Figures 5c and 5d we see that both ϵ and γ are situated in intervals with either lower or higher values. ϵ, which determines how fast the inverse of the market activity reverts to its mean reversion level, is either situated in the interval $(10, 20)$ or in the interval $(60, 80)$. Judging from Figure 5c it seems that ϵ exhibits a regime shifting behaviour. Comparing Figures 5b and 5c we see that ϵ jumps to the lower level when there is a sharp increase in the market activity, and that ϵ reverts to the higher interval when the market activity returns to more normal levels. The following simultaneous behaviour can be observed by inspection of Figures 5c, 5d, 6a and 6b. When the instantaneous volatility

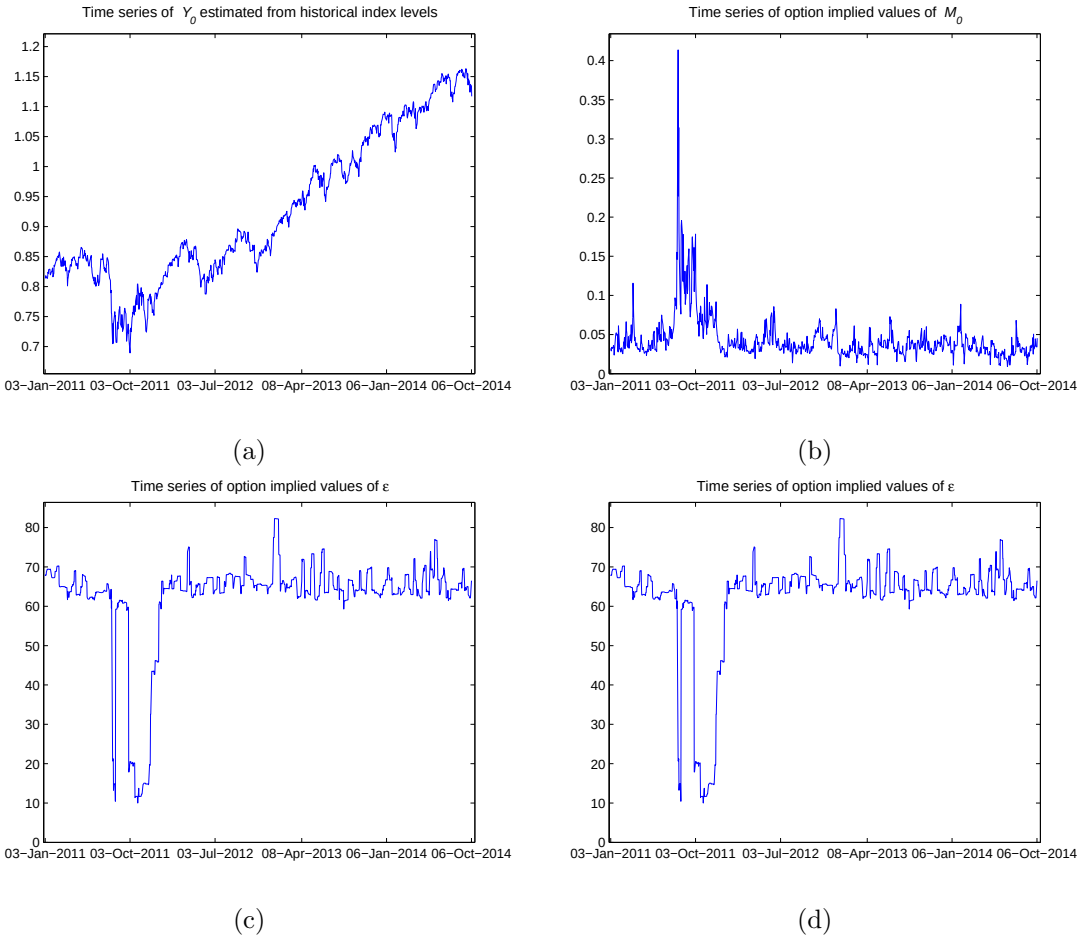


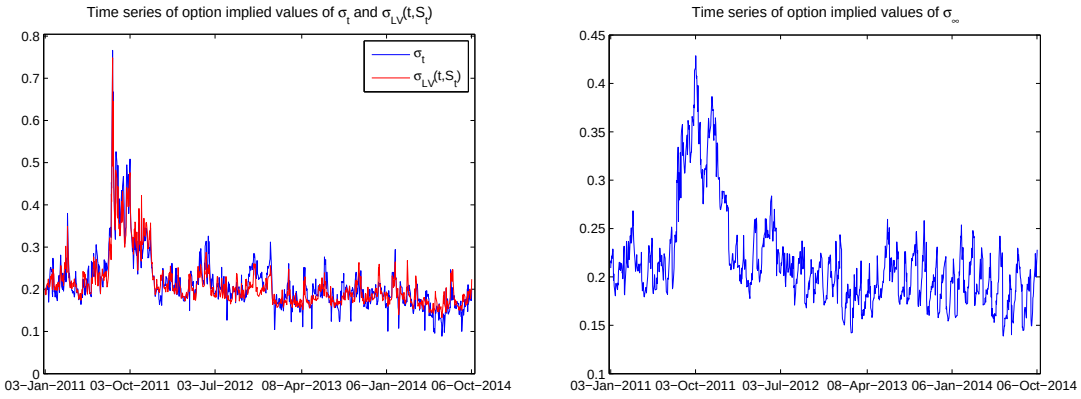
FIGURE 5. Figure (a) shows the time series of the parameter Y_0 estimated from observed index data. Figures (b)-(d) show the time series of option implied values for the parameters M_0, ϵ and γ between 3 January 2011 to 7 October 2014.

increases sharply in Figure 6a, the value of γ in Figure 5d decreases in order to increase the value of the long term average value of the volatility, σ_∞ shown in Figure 6b, and simultaneously the speed of mean reversion, ϵ, decreases in Figure 5c in order to be able to match the implied volatility level corresponding to 91 days to maturity.

As we will see later on, the instantaneous local volatility is a by-product of the estimated option implied volatility surface. In Figure 6a, the instantaneous local volatility $\sigma_{\mathrm{LV}}(t, S_t)$, as defined below, is plotted together with the instantaneous MMM volatility. From the plot we see that the MMM volatility largely follows the instantaneous local volatility. Derman and Kani (1998) showed that under classical no-arbitrage assumptions for local volatility models the squared local volatility is the following expected value under the formally applied risk-neutral measure $\mathbb{Q}$

$$\sigma^2_{\mathrm{LV}}(t, K) = \lim_{T \to t} \mathbb{E}^{\mathbb{Q}} \left[\tilde{\sigma}^2_T | \mathcal{F}_t, S_T = K \right] = \tilde{\sigma}^2_t, \tag{12}$$

where $\tilde{\sigma}_t$ is the instantaneous stochastic volatility for the index and $\mathcal{F}_t$ denotes the information available at time t. As shown in Platen and Heath (2006), for short terms to maturity the benchmark approach gives similar prices as the formally applied risk-neutral



(a) (b)

FIGURE 6. Figure (a) compares the instantaneous local volatility, $\sigma_{\text{LV}}(t, S_t)$, (red) and the option implied instantaneous MMM volatility, σ_t, (blue) between 3 January 2011 to 7 October 2014. Figure (b) shows the evolution of the long term average volatility, σ_∞, over the same period.

approach. Therefore, Figure 6a seems to indicate that also more generally we have that the instantaneous local volatility and the stochastic volatility coincide, i.e.

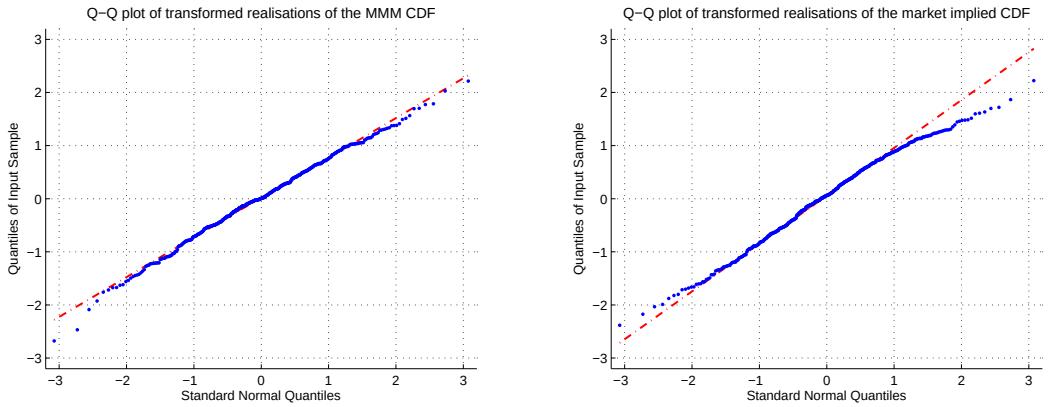
$$\sigma_{\text{LV}}(t, S_t) = \tilde{\sigma}_t. \tag{13}$$

2.4. **Option implied PDFs.** Before we turn to the issue of modeling and measuring the liquidity spreads embedded in option prices, we will shortly analyze whether the time series of estimated MMM parameters are able to explain the realized characteristics of the underlying index over the time period. For each date, t, in the time series we use simulation, based on the estimated MMM parameters, in order to estimate the kernel density $\hat{f}_t$ for the index level two days forward using the method to be described in Section 4.3. Given that the kernel density is the true density for the index, we have that the cumulative distribution function (CDF) of the kernel density evaluated at the actual outcomes for the index level is uniformly distributed. We let $\hat{F}_t$ denote the two day CDF that corresponds to the kernel density $\hat{f}_t$. In order to test if the estimated CDFs can explain the movements of the underlying index, we then create the time series of CDF values evaluated at the actual outcomes for the index, S_{t+2}, two days after the CDF was estimated. Thus, we have that the realized CDF values are given by

$$\hat{F}_t(S_{t+2}) = \int_0^{S_{t+2}} \hat{f}_t(x)dx. \tag{14}$$

Since we want to avoid to introduce dependence between the empirically evaluated CDF values, we evaluate the two day CDFs every second day in the sample. This means that the sample size for the evaluation decreases from 947 to 473. If the estimated CDFs would represent the actual distribution of the underlying index we would have that

$$\hat{F}_t(S_{t+2}) \sim U(0, 1). \tag{15}$$



(a) (b)

FIGURE 7. Figure (a) shows the Q-Q plot of the transformed realized CDF values calculated from kernel densities implied by the MMM parameters for each date in the time series. Figure (b) shows the corresponding Q-Q plot using the CDF calculated directly from the real-world PDF recovered from option data.

Using the inverse transform method we can then transform these uniformly distributed variables to $N(0,1)$ variables according to

$$\Phi^{-1}(\hat{F}_t(S_{t+2})) \sim N(0,1), \tag{16}$$

where Φ is the CDF for a $N(0,1)$ distributed random variable. Hence, if the estimated MMM parameters implied CDFs, $\hat{F}_t$, are such that the empirically evaluated inverse CDFs in equation (16) over the estimation period are approximately $N(0,1)$ distributed, then we cannot reject the hypothesis that the option implied parameter values describe the actual characteristics of the underlying index.

We have also performed the same analysis directly, using the real-world PDFs implied by the observed option prices over the estimation period. The Q-Q plots of the transformed realized CDF values implied by the MMM kernel densities and the market implied real-world PDFs are given in Figures 7a and 7b. In order to be able to draw conclusions from the plots in Figures 7a and 7b we show also a Q-Q plot of the standardized returns of the index in Figure 8. Comparing the Q-Q plots in Figures 7a and 7b with the Q-Q plot of the standardized returns in Figure 8, we draw the conclusion that the market implied real-world density, and especially the real-world density implied by the calibrated MMM parameters, are able to predict quite well the realized characteristics of the underlying S&P 500 index. From Figure 8 we observe the well-known characteristic that the standardized realized index returns over the period exhibit much fatter tails than the normal distribution. From the Q-Q plot in Figure 7b we draw the conclusion that the market implied real-world density can explain the fat tails in the realized return data, but the plot also indicates that the market implied density predicts fatter tails than what is realized by the index. In the next subsection we will argue that this is consistent with the existence of a liquidity premium embedded in index options with strikes far from the current index level (see e.g. Gârleanu *et al.*, 2009). Inspecting the Q-Q plot in Figure 7a we conclude that also the MMM implied PDF has the ability to explain the fat tails in the realized returns. Since the MMM has been shown to be able to accurately describe the

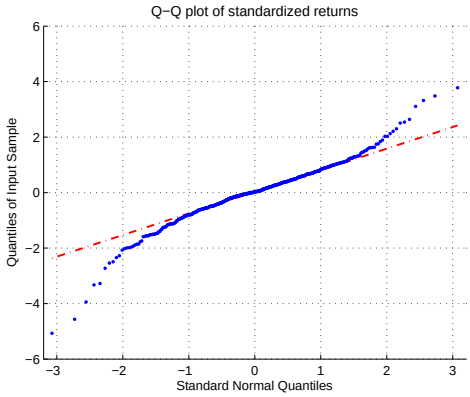


FIGURE 8. Q-Q plot of the standardized returns of the S&P 500 index over the period 3 January 2011 to 7 October 2014.

TABLE 2. p-values and outcomes for the three statistical tests for the empirically evaluated inverse MMM CDFs in equation (16) over the period 3 January 2011 to 7 October 2014.

Test	Jarque-Bera	Kolmogorov-Smirnov	Lilliefors
p-value	0.137	0.659	0.239
Null hypothesis	not rejected	not rejected	not rejected

characteristics of broad equity indices, the estimated MMM parameters do, however, not seem to include the liquidity premium in the tails which is present in the option implied densities (see Figure 7b). From Figure 7a we see that the MMM PDF has removed the liquidity premia in the tails of the PDF, and visually seems to be very well able to predict the realized statistical characteristics of the index. Since the transformed realized CDF values implied by the MMM approximately lie on a straight line, we draw the conclusion that the MMM PDF can explain the shape of the realized return distribution. However, since the slope of the line is less than one, the volatility implied from the options by the MMM density is larger than what is realized by the index.

We have also used three different statistical tests in order to test the null hypothesis that the empirically evaluated inverse MMM CDFs in equation (16) come from a normal distribution, against the alternative that they do not come from such a distribution. The three statistical tests that we use are Jarque-Bera (Jarque and Bera, 1981), Kolmogorov-Smirnov (see e.g. Stephens, 1974) and Lilliefors (Lilliefors, 1967). The significance level for all three tests was 5%, meaning that the probability of falsely rejecting the null hypothesis when it is true is 5%. Since the standard deviation of the transformed realized CDF values is less than one, we have used the values divided by the standard deviation as input to the Kolmogorov-Smirnov test. The p-values and the outcomes of the three statistical tests are given in Table 2. From the table we see that the null hypothesis that the empirically evaluated inverse MMM CDFs come from a normal distribution cannot be rejected at the 5 % level. Given our analysis we conclude that the MMM seems to exhibit properties which lets us both recover the real-world PDF and liquidity premia from observed S&P 500 index options.

TABLE 3. Time-to-maturity and moneyness grid.

Time-to-maturity	5 days	30 days	91 days	1 year
Moneyness interval	[0.95, 1.03]	[0.80, 1.10]	[0.75, 1.15]	[0.70, 1.25]

2.5. **Discovering the liquidity premium in option prices.** As is apparent from the PDF and implied volatility graphs in Figure 4, there seems to be a discrepancy between the implied volatilities given by the MMM and the market via the option prices, respectively. The MMM implied volatility appears to be lower than what is quoted in the market, which means that the option prices in the market are higher than or equal to those obtained from the calibrated MMM. When studying the time series of the spread between the MMM and market implied volatility, it is revealed that the market prices of options with longer maturities are higher than the corresponding MMM prices. When studying the time series of implied volatility spreads we also identify a positive spread for options with low strike prices, i.e. for OTM puts. This is consistent with the recent study by Gârleanu *et al.* (2009), where the authors identify a demand based premium for quoted market prices of S&P 500 options, and especially for OTM puts. It is well documented that there is excess demand for OTM puts, which are used by institutions to hedge US equity portfolios. It seems that this excess demand translates into higher prices for those options. The spread can additionally be explained by the fact that liquidity is lower for longer term options and options with strikes far from the current index level. Hence the difference in the prices of options implied by the MMM and quoted in the market can in part be explained by what we call here for simplicity a liquidity spread.

In order to model and estimate the liquidity spread in implied volatilities we first create the time series of spreads between the MMM and market implied volatilities for points on a non-uniform moneyness and time-to-maturity grid that lie in the intervals given in Table 3. For each time-to-maturity we use 15 evenly distributed grid points between the limits of the intervals. We thus have 60 grid points with different time-to-maturity and moneyness, $\{\tau_i, m_i\}_{i=1}^{60}$, where τ_i is the time-to-maturity for grid point i and m_i is the corresponding moneyness. The observed spreads in implied volatilities on 3 October 2014, are illustrated in Figure 9. This illustration of the spreads is typical for the observed spreads for most days in the full sample. In order to be able to mimic the shape in the figure we propose the following parametric model for the liquidity spread in implied volatilities

$$Y_{t,i} = \alpha_0 + \alpha_1 \tau_i + \mathrm{e}^{a\tau_i}\alpha_2 (m_i - 1)^2 \mathbb{1}_{\{m_i \leq 1\}} + \mathrm{e}^{b\tau_i}\alpha_3 (m_i - 1)^2 \mathbb{1}_{\{m_i > 1\}} + \epsilon_{t,i}, \tag{17}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function, $Y_{t,i}$ is the at time t observed spread for grid point i and $\epsilon_{t,i}$ is an error term. The parameter α_1, thus, provides a measure of the term liquidity premium, whereas the parameters α_2 and α_3 measure the liquidity premia in the left and right tail, respectively. We use a quadratic function to model the liquidity spread in the tails motivated by the shape of the observed spreads illustrated in Figure 9. In order to facilitate a decreasing curvature for the tail spread for longer maturities we have included the exponential terms $\mathrm{e}^{a\tau_i}$ and $\mathrm{e}^{b\tau_i}$. The model for the liquidity spread also includes the constant α_0.

2.5.1. *Estimation results.* For each date t in the data set, we fit the parametric model (17) to the observed spreads $\{Y_{t,i}\}_{i=1}^{60}$ with linear least squares, where $\sum_{i=1}^{60} \epsilon_{t,i}^2$ is minimized. For convenience we have used $a = -\ln(2)$ and $b = -\ln(2)$ in equation (17) when estimating

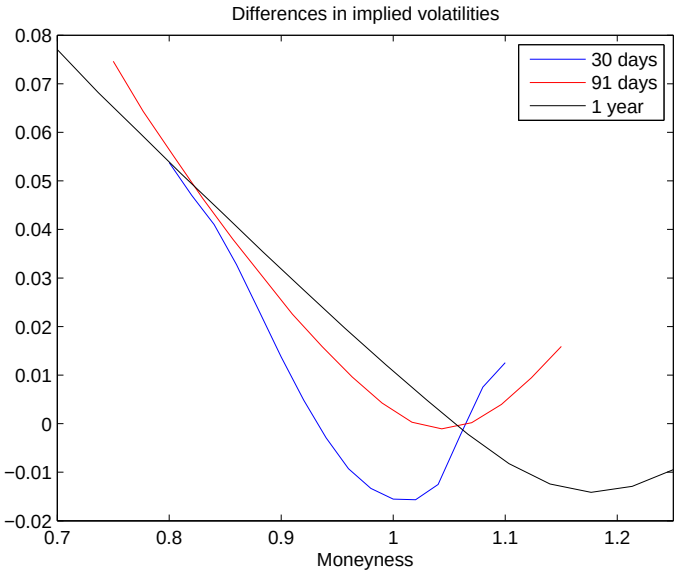


FIGURE 9. Difference in implied volatilities between the market and implied by the estimated MMM parameters on 3 October 2014. The blue, red and black curves correspond to options with 30 days, 91 days and 1 year to maturity respectively.

the parameters $\alpha_0, \alpha_1, \alpha_2$ and α_3. In this way we obtain four time series of parameter estimates $\{\hat{\alpha}_0(t), \hat{\alpha}_1(t), \hat{\alpha}_2(t), \hat{\alpha}_3(t)\}$ over the period from 3 January 2011 to 7 October 2014. For the short period during 2011, when the estimated values for ϵ and γ lay in the lower interval, the spreads in implied volatilities were small and for some dates even negative. For all other dates in the estimation period, the market implied volatilities for low strikes and for longer maturities were consistently higher than the corresponding MMM implied volatilities. In order to estimate the properties of $\hat{\alpha}_i(t)$ we use Wilcoxons signed-rank test for each time series. Under the null hypothesis that the observations $\{\hat{\alpha}_i(t)\}$ come from a continuous symmetric distribution, the procedure provides us with parameter estimates as well as confidence intervals for the parameter estimates. The estimated parameter values together with the 95% confidence intervals are given in Table 4. The table also contains the p-values from Wilcoxons signed-rank test for each of the parameters. From the p-values we conclude that the null hypothesis cannot be rejected at the 5% level for α_0, α_1 and α_3. For α_2 however, the null hypothesis is rejected at the 5% level. For this reason we have not included a confidence interval for α_2 in Table 4. Inspecting the histogram of the parameter estimates $\{\hat{\alpha}_2(t)\}$ in Figure 10, we observe that the values are not symmetrically distributed around the median. The procedure should however provide us with a reasonable estimate of α_2 since most of the observations are centered around the estimated value 1.65.

As we can see from the parameter estimates in Table 4, where $\alpha_2 > \alpha_3$, the liquidity spread in the left tail is much more pronounced than for the right tail, where the estimated spread is even slightly negative. From Table 4 we also see that the term spread parameter implies a liquidity spread for one year ATM options of $\alpha_0 + \alpha_1 = 2.61\%$. Figures 11a and 11b illustrate the estimated liquidity spreads for maturity dates which are 91 days and one year into the future, respectively. As is apparent from the figures we have estimated

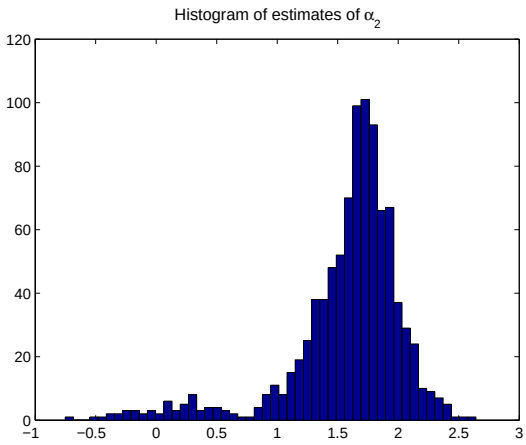


FIGURE 10. Histogram of the estimates $\{\hat{\alpha}_2(t)\}$ of the parameter α_2 in the parametric model (17) over the period 3 January 2011 to 7 October 2014.

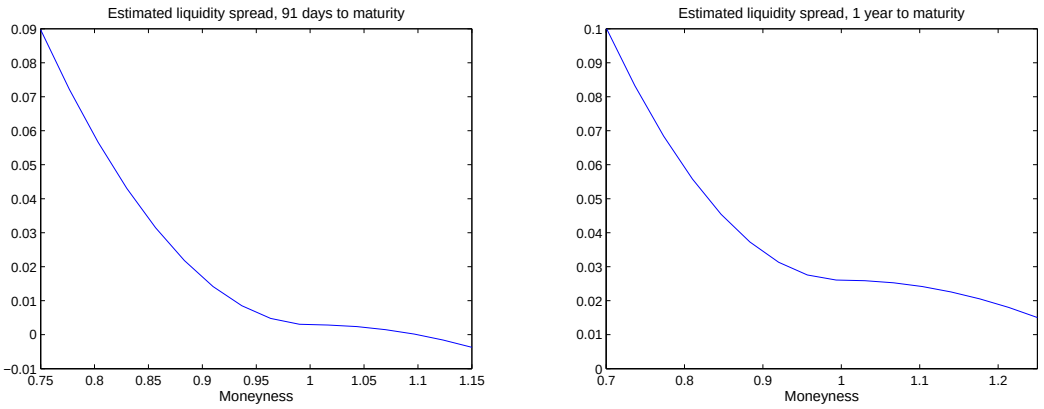
TABLE 4. Parameter estimates for the parametric model (17) for the full sample.

Parameter	α_0	α_1	α_2	α_3
Estimate	-0.0047	0.0308	1.65	-0.352
	(-0.0052, -0.0042)	(0.0295, 0.0321)	(-)	(-0.373, -0.331)
p-value	0.842	0.818	0.032	0.338

a substantial liquidity spread for options with low strikes, i.e. for OTM puts that are typically in high demand. The discovered liquidity premia are consistent with what has earlier been found in the literature (see e.g. Bollen and Whaley, 2004; Gârleanu *et al.*, 2009). From Figure 11b we also see that we have discovered a term spread for ATM options. This observation is consistent with the findings in Platen (2009).

3. General modeling framework

In this section we will describe the general modeling framework that we use in order to simultaneously recover the real-world density and liquidity premia from option data. We will first show how to price derivatives using the real-world probability measure under the benchmark approach. The real-world pricing formula relies on the existence of a strictly positive numéraire portfolio that when used as benchmark makes all benchmarked nonnegative portfolios supermartingales. The critical assumption of the paper is that the numéraire portfolio can be approximated by a well-diversified portfolio, e.g. the S&P 500 total return index when the investment universe represents the US equity market. In Platen (2005) it is shown that diversified portfolios can be used as an approximation of the numéraire portfolio, which justifies the use of diversified equity indices as proxies for the numéraire portfolio in derivative pricing. The benchmark approach has recently been applied in the literature to real-world pricing of currency derivatives using a 3/2 stochastic volatility model (see Baldeaux *et al.*, 2015). We formulate below the notion of strong arbitrage which is ruled out by the modeling framework that we present. This arbitrage definition is similar to the arbitrage formulation given in Loewenstein and Willard (2000). In the second part of this section we will summarize some properties of the generalized



(a) (b)

FIGURE 11. The liquidity spreads in implied volatilities fitted by the parametric model (17) over the period 3 January 2011 to 7 October 2014. Figure (a) shows the spread in implied volatilities for options with 91 days to maturity and Figure (b) shows the corresponding plot for options with 1 year to maturity.

MMM that we use as a tool in order to estimate the above documented liquidity premia embedded in equity index option prices.

3.1. **Real-world pricing under the benchmark approach.** The summary of the benchmark approach that we present here is based on results presented in Platen (2009) and Platen (2011). We let $\tilde{S}_t$ denote the S&P 500 total return index, which is according to Platen and Heath (2006) an approximation for the numéraire portfolio for the US equity market. Thus, since we assume a deterministic instantaneous dividend yield $\{\delta_t\}_{t\geq 0}$, we have that the S&P 500 price index is given by

$$S_t = \mathrm{e}^{-\int_0^t \delta_s ds} \tilde{S}_t. \tag{18}$$

Assuming a deterministic instantaneous interest rate, $\{r_t\}_{t\geq 0}$, we denote by

$$B_t = \mathrm{e}^{\int_0^t r_s ds}, \tag{19}$$

the value of the savings account at time t.

We let P_t^h denote the portfolio value at time t of an arbitrary self-financing portfolio, where h represents the predictable vector of fractions invested in the primary security accounts. Furthermore, $\mathcal{V}^+$ denotes the set of all strictly positive, finite, and self-financing portfolios with initial capital $x > 0$. We use the following definition, formulated in Platen (2011), for the numéraire portfolio.

Definition 3.1. *For a given $x > 0$, a strictly positive, finite, self-financing portfolio $P_t^{h^*} \in \mathcal{V}^+$ is called a numéraire portfolio if all non-negative portfolios P_t^h, when denominated in units of $P_t^{h^*}$, are supermartingales.*

The supermartingale property yields the uniqueness of the value process of the numéraire portfolio (see Platen, 2011). When we price derivatives under the benchmark approach we choose the numéraire portfolio as numéraire or benchmark. We let $\hat{P}_t^h$ denote the

benchmarked value of the portfolio P_t^h and hence we have that

$$\hat{P}_t^h = \frac{P_t^h}{P_t^{h^*}}. \tag{20}$$

From Definition 3.1 we have that the benchmarked value $\hat{P}_t^h$ of any non-negative portfolio P_t^h satisfies the supermartingale property

$$\hat{P}_t^h \geq \mathbb{E}^{\mathbb{P}}\left(\hat{P}_s^h \middle| \mathcal{F}_t\right), \tag{21}$$

for all $0 \leq t \leq s \leq \infty$. Note that we denote by $\mathbb{P}$ the real-world probability and by $\mathbb{E}^{\mathbb{P}}$ the real-world expectation. For the case where we have equality in relation (21) the benchmarked portfolio price forms a martingale, in which case we call the price *fair*. There can be many supermartingales that approach the same future random value. The martingale that matches the future random value is the minimal possible supermartingale. This leads to the following *Law of the Minimal Price* given in Platen (2009).

Theorem 3.1. (Law of the Minimal Price) *If a fair portfolio replicates a given non-negative pay-off at a bounded stopping time, then this portfolio represents the minimal replicating portfolio among all non-negative portfolios that replicate this pay-off.*

Since there may exist self-financing portfolios in a benchmark setting that are not fair, the classical Law of one Price does not hold in general. For a given hedgeable pay-off the corresponding fair hedge portfolio represents the least expensive hedge portfolio.

We now turn to the problem of pricing a particular pay-off. We let H_T denote a contingent claim that delivers a non-negative pay-off at maturity T. The benchmarked pay-off

$$\hat{H}_T = \frac{H_T}{P_T^{h^*}}, \tag{22}$$

is assumed to have finite expectation,

$$\mathbb{E}^{\mathbb{P}}\left(\hat{H}_T \middle| \mathcal{F}_0\right) < \infty. \tag{23}$$

As formulated in Platen (2011), the fair price of the contingent claim H_T is by the Law of the Minimal Price given by the *real-world pricing formula*:

Corollary 3.1. *If for a contingent claim H_T, there exists a fair self-financing portfolio $P_t^{h_H}$ that replicates this claim at maturity T such that $H_T = P_T^{h_H}$, then its minimal replicating price at time $t \in [0, T]$ is given by the real-world pricing formula*

$$P_t^{h_H} = P_t^{h^*} \mathbb{E}^{\mathbb{P}}\left(\frac{H_T}{P_T^{h^*}} \middle| \mathcal{F}_t\right). \tag{24}$$

We will now illustrate that risk-neutral pricing is a special case of real-world pricing. For simplicity, we will illustrate the connection between risk-neutral and real-world pricing for $t = 0$. By introducing the process $\Lambda = \{\Lambda_t, t \geq 0\}$ for the normalized benchmarked savings account,

$$\Lambda_t = \frac{\hat{B}_t}{\hat{B}_0} = \frac{B_t / P_t^{h^*}}{B_0 / P_0^{h^*}}, \tag{25}$$

we can, since $B_0 = 1$ and $\Lambda_0 = 1$, express the real-world pricing formula as

$$P_0^{h_H} = \mathbb{E}^{\mathbb{P}}\left(\Lambda_T \frac{H_T}{B_T}\middle|\mathcal{F}_0\right). \tag{26}$$

In our market, we identify Λ_t as the Radon-Nikodym derivative process for the formally obtained risk-neutral measure (see Platen and Heath, 2006) and we have that

$$\mathbb{E}^{\mathbb{P}}\left(\Lambda_T\middle|\mathcal{F}_0\right) \leq \Lambda_0 = 1, \tag{27}$$

since Λ is a supermartingale. Λ can be interpreted as the nonnegative local martingale that would determine the respective risk-neutral measure if such an equivalent measure would exist. The formally taken risk-neutral price $\tilde{P}_0^{h_H}$ of the contingent claim is given by (see e.g. Björk, 2011)

$$\tilde{P}_0^{h_H} = \frac{\mathbb{E}^{\mathbb{P}}\left(\Lambda_T \frac{H_T}{B_T}\middle|\mathcal{F}_0\right)}{\mathbb{E}^{\mathbb{P}}\left(\Lambda_T\middle|\mathcal{F}_0\right)}, \tag{28}$$

and since Λ is a supermartingale we have that,

$$P_0^{h_H} \leq \tilde{P}_0^{h_H}. \tag{29}$$

The conclusion is that only if the Radon-Nikodym derivative Λ is a martingale, then the real-world price and the risk-neutral price are guaranteed to be equivalent. Otherwise, the real-world price is lower than or equal to the risk-neutral price. Platen (2009) argues, through an empirical example for the S&P 500 total return index, that the typical path of a normalized benchmarked savings account seems highly unlikely to be the realization of a martingale, and rather resembles the trajectory of a strict supermartingale. The conclusion from this is that the risk-neutral prices of derivatives can be substantially higher than the corresponding real-world prices, for longer term contracts. Since the market standard is to use risk-neutral pricing in order to price derivatives, the benchmark framework enables the identification of the term premium embedded in index option prices.

Under the benchmark approach, the notion of arbitrage opportunities will be different from the classical risk-neutral approach. Platen (2011) argues that the economically meaningful form of arbitrage opportunities that must be excluded by a model is so called *strong arbitrage*. Strong arbitrage arises when a market participant can generate with a nonnegative portfolio strictly positive wealth from zero initial capital. The definition of strong arbitrage, motivated by the supermartingale property of benchmarked prices, was introduced in Platen (2002) and is given below:

Definition 3.2. *A nonnegative portfolio P^h is a strong arbitrage if it starts with zero initial capital, that is $P_0^h = 0$, and generates strictly positive wealth with strictly positive probability at a later time $t \in (0, \infty)$, that is, $\mathbb{P}(P_t^h > 0) > 0$.*

A similar definition of arbitrage was formulated also in Loewenstein and Willard (2000) based on economic reasoning. As argued by Platen (2011), weaker forms of classical arbitrage may still exist, but this does, however, not harm the economic viability of the market model. By exploiting the supermartingale property of benchmarked prices it is shown in Platen (2011) that there does not exist any nonnegative portfolio that is a strong arbitrage under the given benchmark framework.

3.2. **Pricing of S&P 500 index derivatives under the benchmark approach.** In our approximate setting, where the S&P 500 total return index $\tilde{S}_T = S_T \mathrm{e}^{\int_0^T \delta_s ds}$ is assumed to approximate the numéraire portfolio for the US equity market, we will now give the expressions for fair prices of European call and put options, futures and zero coupon bonds under the benchmark approach. The real-world price of a European call option, $C(K,T)$, with strike price K and maturity T, written on the S&P 500 index, is given by

$$C(K,T) = \tilde{S}_t \mathbb{E}^{\mathbb{P}} \left(\frac{(S_T - K)^+}{\tilde{S}_T} \middle| \mathcal{F}_t \right) = \tilde{S}_t \mathbb{E}^{\mathbb{P}} \left(\left(\mathrm{e}^{-\int_t^T \delta_s ds} - \frac{K}{\tilde{S}_T} \right)^+ \middle| \mathcal{F}_t \right). \tag{30}$$

Similarly, the real-world price of the put option, $P(K,T)$, with the same strike and maturity is under the benchmark approach given by

$$P(K,T) = \tilde{S}_t \mathbb{E}^{\mathbb{P}} \left(\frac{(K - S_T)^+}{\tilde{S}_T} \middle| \mathcal{F}_t \right) = \tilde{S}_t \mathbb{E}^{\mathbb{P}} \left(\left(\frac{K}{\tilde{S}_T} - \mathrm{e}^{-\int_t^T \delta_s ds} \right)^+ \middle| \mathcal{F}_t \right). \tag{31}$$

The real-world futures price is given by solving for the futures price $F_{t,T}$ in the equation

$$\tilde{S}_t \mathbb{E}^{\mathbb{P}} \left(\frac{S_T - F_{t,T}}{\tilde{S}_T} \middle| \mathcal{F}_t \right) = 0, \tag{32}$$

and thus the fair futures price is for a deterministic dividend rate given by

$$F_{t,T} = \frac{\mathbb{E}^{\mathbb{P}} \left(\frac{S_T}{\tilde{S}_T} \middle| \mathcal{F}_t \right)}{\mathbb{E}^{\mathbb{P}} \left(\frac{1}{\tilde{S}_T} \middle| \mathcal{F}_t \right)} = \frac{1}{\mathbb{E}^{\mathbb{P}} \left(\frac{1}{S_T} \middle| \mathcal{F}_t \right)}. \tag{33}$$

The benchmarked fair price of a zero coupon bond, $P_{t,T}$, is given by the conditional expectation of the benchmarked value of a contract that pays out one monetary unit at maturity, i.e. $H_T = 1$. Thus, we have

$$P_{t,T} = \tilde{S}_t \mathbb{E}^{\mathbb{P}} \left(\frac{1}{\tilde{S}_T} \middle| \mathcal{F}_t \right). \tag{34}$$

Using the payout function $H_T = 1$ in equation (28), and assuming evaluation at time 0, we arrive after some calculation at the following expression for the fair price of a zero coupon bond

$$P_{0,T} = \frac{1}{B_T} \mathbb{E}^{\mathbb{P}} \left(\Lambda_T \middle| \mathcal{F}_0 \right), \tag{35}$$

where we identify

$$\tilde{P}_{0,T} = \frac{1}{B_T} \tag{36}$$

as the formally taken risk-neutral price of a zero coupon bond. Since $\mathbb{E}^{\mathbb{P}} \left(\Lambda_T \middle| \mathcal{F}_0 \right) \leq 1$ this means that

$$P_{0,T} \leq \tilde{P}_{0,T}, \tag{37}$$

and hence we have that the fair zero coupon bond price is smaller than or equal to the corresponding risk-neutral price. Similarly, it can be shown that the fair futures price, $F_{0,T}$, is greater than or equal to the corresponding risk-neutral price, $\tilde{F}_{0,T}$,

$$F_{0,T} \geq \tilde{F}_{0,T}. \tag{38}$$

Since fair prices of options, futures and bonds under the benchmark approach differ from the corresponding risk-neutral prices, we have to be careful when we extract implied volatilities from options priced under the real-world probability measure. We will, therefore, formulate here put-call parity under the benchmark approach in order to illustrate how implied volatilities that are consistent for both call and put options can be extracted in this framework. Studying the difference between real-world prices of call and put options with strike price K and maturity T yields

$$
\begin{aligned}
&C(K,T) - P(K,T) = \\
&\tilde{S}_t \mathbb{E}^{\mathbb{P}} \left(\left(\mathrm{e}^{-\int_t^T \delta_s ds} - \frac{K}{\tilde{S}_T} \right)^+ \middle| \mathcal{F}_t \right) - \tilde{S}_t \mathbb{E}^{\mathbb{P}} \left(\left(\frac{K}{\tilde{S}_T} - \mathrm{e}^{-\int_t^T \delta_s ds} \right)^+ \middle| \mathcal{F}_t \right) = \\
&\tilde{S}_t \mathbb{E}^{\mathbb{P}} \left(\left(\mathrm{e}^{-\int_t^T \delta_s ds} - \frac{K}{\tilde{S}_T} \right)^+ - \left(\frac{K}{\tilde{S}_T} - \mathrm{e}^{-\int_t^T \delta_s ds} \right)^+ \middle| \mathcal{F}_t \right) = \\
&\tilde{S}_t \mathbb{E}^{\mathbb{P}} \left(\mathrm{e}^{-\int_t^T \delta_s ds} - \frac{K}{\tilde{S}_T} \middle| \mathcal{F}_t \right) = \tilde{S}_t \mathrm{e}^{-\int_t^T \delta_s ds} - K \tilde{S}_t \mathbb{E}^{\mathbb{P}} \left(\frac{1}{\tilde{S}_T} \middle| \mathcal{F}_t \right) = \\
&\tilde{S}_t \mathrm{e}^{-\int_t^T \delta_s ds} - K P_{t,T}. \qquad (39)
\end{aligned}
$$

Thus, this expression is analogous to the put-call parity under risk-neutral pricing but where the risk-neutral price of a zero coupon bond is replaced by the fair zero coupon bond price. In order to get the same value of Black-Scholes implied volatilities for call and for put option prices estimated under the benchmark approach, we need to use the interest rate over the period until maturity implied by the fair zero-coupon bond rather than the risk-neutral zero-coupon bond price observed on the market.

3.3. **Further properties of the generalized Minimal Market Model.** By the generalized MMM, as introduced in Section 2.2, we have that the annualized squared instantaneous volatility of the returns of the normalized index process Y in t-time is given by

$$\sigma_t^2 = \frac{M_t}{Y_{\tau_t}}. \qquad (40)$$

Application of the Itô formula, using equations (3), (4), (5) and (6), yields the SDE for the discounted index level $\bar{S}_t$

$$d\bar{S}_t = \bar{S}_t M_t \left(a - 1 + \frac{1}{Y_{\tau_t}} \right) dt + \bar{S}_t \sigma_t dW_t = \bar{S}_t (\mu_t dt + \sigma_t dW_t), \qquad (41)$$

with initial value $\bar{S}_0 = A_0 Y_0$, and where by (40) the instantaneous expected rate of return is given by

$$\mu_t = M_t \left(a - 1 + \frac{1}{Y_{\tau_t}} \right) = \sigma_t^2 + M_t (a - 1). \qquad (42)$$

Thus, from equation (42) we have that the risk premium of the index is given by $\sigma_t^2 + M_t(a-1)$.

The savings account, when benchmarked with the total return index $\tilde{S}_t$, that is

$$\hat{B}_t \equiv \frac{B_t}{\tilde{S}_t} = \frac{1}{\bar{S}_t}, \qquad (43)$$

is the inverse of the discounted total return index value. From equation (41) and the Itô formula we have that $\hat{B}_t$ is characterized by the SDE

$$d\hat{B}_t = \hat{B}_t(\sigma_t^2 - \mu_t)dt - \sigma_t dW_t, \tag{44}$$

for $t \geq 0$. From equation (44) it follows that if

$$\sigma_t^2 \leq \mu_t \tag{45}$$

for all $t \geq 0$, then $\hat{B}_t$ forms an $(\mathcal{F}, \mathbb{P})$-supermartingale, i.e.

$$\hat{B}_t \geq \mathbb{E}^{\mathbb{P}}(\hat{B}_s \mid \mathcal{F}_t), \text{ for } s \geq t. \tag{46}$$

This supermartingale property is the key property of any benchmarked nonnegative security under the benchmark approach. Thus, the possibility of strong arbitrage is eliminated since a nonnegative supermartingale that reaches zero will always remain at zero. Since

$$\mu_t = \sigma_t^2 + M_t(a-1), \tag{47}$$

we must have, since $M_t > 0$, that

$$a \geq 1 \tag{48}$$

in order for the model to be free of strong arbitrage. In this case the long term average net growth rate, a, is high enough to make the benchmarked savings account a supermartingale. As shown in Table 1, the estimate for the parameter a is with 1.44 indeed significantly higher than one.

4. Estimation technique

In this section we will present the estimation technique that we use for estimating the MMM parameters implied by observed prices of S&P 500 index options. First we describe the method that we use in order to produce high quality estimates of the option implied risk-neutral density (RND). The RND will then be used in order to recover the real-world probability density function (PDF) from option prices by using a result in Heath and Platen (2006). In the last part of this section we describe the Simulated Maximum Likelihood algorithm that we use in order to estimate the MMM parameters that are implied by the observed option prices.

4.1. **RND recovery.**

As discussed in Platen and Heath (2006), for short dated derivatives with up to three years time-to-maturity and similar to those we consider, the quantitative difference between real-world pricing under the benchmark approach and the formally applied risk-neutral pricing are rather small. Therefore, we can recover the formally obtained RND from market prices of options, similar to current market practise. We formulate an optimization model that has the squared local volatility (LV) as the decision variable. Estimation of the local volatility surface (LVS) is an ill-posed inverse problem, which has gained a lot of attention in the literature. There are many examples of papers in the literature which describe the estimation of local volatility surfaces from observed option prices (see e.g. Andersen and Brotherton-Ratcliffe, 1997; Lagnado and Osher, 1997; Bouchouev and Isakov, 1999; Coleman *et al.*, 1999). Most of these methods do, however, not satisfactorily balance the objectives of being consistent with market prices while producing surfaces that are realistic and stable over time. One of the reasons that we choose LV as decision variable is that it does not involve an integration of other variables. This makes it much easier to separate the noise in the estimation. We estimate high-quality local volatility surfaces with a non-parametric method developed in Barkhagen and Blomvall

(2015), which produces realistic surfaces that are consistent with observed options and futures prices. The RND recovery is then a by-product of the estimation problem that we solve.

The local volatility estimation framework that we use should be seen as a tool for producing option implied volatility surfaces of high quality. Thus, we do not restrict the model for the index to be given by a local volatility function model for recovery of the RND. The estimation framework can be better seen as an advanced interpolation tool for producing a smooth and plausible call price surface.

When setting up the problem for the estimation of the implied LVS, we start with the Black-Scholes-Merton PDE that describes the dynamics of the option price, $V(t,S)$, in the local volatility framework

$$\frac{\partial V(t,S)}{\partial t} + (r_t - \delta_t)S\frac{\partial V(t,S)}{\partial S} + \tfrac{1}{2}\sigma^2(t,S)S^2\frac{\partial^2 V(t,S)}{\partial S^2} = r_t V(t,S), \tag{49}$$

where the interest rate, r_t, and dividend yield, δ_t, are assumed to be deterministic functions of time. In order to estimate the unknown squared local volatility function, $\sigma^2(t,S)$, the PDE (49) is discretized, which gives rise to a finite difference scheme. In order to increase the efficiency of the finite difference discretization we follow Andersen and Brotherton-Ratcliffe (1997), and change variables in (49) to $x = \ln(S)$ and $H(t,x) = V(t,S)$, giving rise to the PDE,

$$\frac{\partial H(t,x)}{\partial t} + b(t,x)\frac{\partial H(t,x)}{\partial x} + \tfrac{1}{2}v(t,x)\frac{\partial^2 H(t,x)}{\partial x^2} = r_t H(t,x), \tag{50}$$

where $x = \ln(S)$, $b(t,x) = r_t - \delta_t - \frac{1}{2}v(t,x)$ and $v(t,x) = \sigma^2(t,\mathrm{e}^x)$. The PDE (50) is discretized and solved on a fixed grid, where the (t,x) plane is divided into a non-uniformly spaced mesh with $M+1$ nodes along the t axis and $N+2$ nodes along the x axis

$$t_i = t_0 + \sum_{k=0}^{i-1}\Delta_t(k), \quad i = 0,\ldots,M, \tag{51}$$

$$x_j = x_0 + \sum_{k=0}^{j-1}\Delta_x(k), \quad j = 0,\ldots,N+1, \tag{52}$$

where $\Delta_t(i) = t_{i+1} - t_i$ and $\Delta_x(j) = x_{j+1} - x_j$. The grid is chosen so that the current index level is included as a point on the grid in the x direction, and so that the maturities of the observed option prices are included in the nodes along the t axis. We let l denote the integer yielding the initial index level such that

$$\mathrm{e}^{x_l} = S_{\text{init}}, \tag{53}$$

where S_{init} is the initial or current index level. In Andersen and Brotherton-Ratcliffe (1997) it is shown that the discretization of (50) for a uniform grid gives rise to a discretized version of the Fokker-Planck equation, which describes the relationship between the discretized LVS and the discrete risk-neutral density (RND) surface. The generic form of the discretized Fokker-Planck equation can be written as

$$\boldsymbol{q}_{i+1}^T = \boldsymbol{q}_i^T f_i(\hat{\boldsymbol{v}}, \boldsymbol{\Gamma}), \quad i = 0,\ldots,M-1, \tag{54}$$

with initial condition $\boldsymbol{q}_0 = \boldsymbol{q}_l$, where $\boldsymbol{q}_l$ is a vector where the lth element is one and all other elements are zero. In equation (54), $\boldsymbol{q}_i \in \mathbb{R}^{N\times 1}$ is the discretized RND at time t_i, and $f_i(\hat{\boldsymbol{v}}, \boldsymbol{\Gamma})$ is a matrix that is a function of the unknown discretized LVS, $\hat{\boldsymbol{v}}$, and dividend

terms $\Gamma_i = \exp\left(\int_0^{t_i} \delta_s ds\right)$, $i = 0, \ldots, M$ (see Andersen and Brotherton-Ratcliffe, 1997, for details).

Let $\mathcal{M}$ denote the set of $n_m = |\mathcal{M}|$ instruments for which mid-market prices exist. The functions $g_k(\boldsymbol{q})$, where $\boldsymbol{q} = \{\boldsymbol{q}_i\}_{i=1}^M$ are given from the LVS by equation (54), are used to transform the discretized RND into prices for the specific derivative instruments. For the case where the implied RND exactly produces the observed market prices, $y_{m,k}$, we have

$$y_{m,k} = g_k(\boldsymbol{q}), \qquad k \in \mathcal{M}, \tag{55}$$

or equivalently

$$\boldsymbol{y}_m = \boldsymbol{g}_m(\boldsymbol{q}), \tag{56}$$

where $\boldsymbol{y}_m^T = \{y_{m,k}\}_{k\in\mathcal{M}}$ and $\boldsymbol{g}_m(\boldsymbol{q})^T = \{g_k(\boldsymbol{q})\}_{k\in\mathcal{M}}$. Since the strikes of the options in the set $\mathcal{M}$ might not be included in the discretized grid we will use an interpolation, based on a second-order Taylor expansion, of option prices on the grid. The presence of noise in price data means that there is an uncertainty regarding the observed prices. The noise stems from market micro structure effects such as the bid-ask spread, discrete price ticks, non-synchronized prices or even misprinted prices. In order to take the uncertainty in observed prices into account, and to satisfy the no-arbitrage conditions, we allow deviations from observed prices by penalizing deviations, $\boldsymbol{z}_m \in \mathbb{R}^{n_m}$, from mid-market prices. We introduce the diagonal matrix $\mathbf{F}_m \in \mathbb{R}^{n_m \times n_m}$ that indicates which instruments are allowed to deviate from observed prices, and the diagonal matrix $\mathbf{E}_m \in \mathbb{R}^{n_m \times n_m}$ contains the penalties for deviating from observed prices. In order to guarantee that $\boldsymbol{q}$ represents a discrete probability distribution we introduce a lower bound equal to zero for the RND, and also in each time step ensure that $\mathbf{1}^T \boldsymbol{q}_i = 1$. The necessary no-arbitrage conditions for the resulting squared local volatility surface are satisfied by requiring that the discretized LVS, $\hat{\boldsymbol{v}}$, is greater than zero.

The estimation problem that we solve in order to estimate a realistic discretized squared local volatility surface $\hat{\boldsymbol{v}}$ and RND surface $\boldsymbol{q}$, which are consistent with observed market prices, is given by

$$\begin{aligned}
\min_{\hat{\boldsymbol{v}},\boldsymbol{q},\boldsymbol{\Gamma},\boldsymbol{z}_m} \quad & h(\hat{\boldsymbol{v}}) + \tfrac{1}{2}\boldsymbol{z}_m^T \mathbf{E}_m \boldsymbol{z}_m \\
\text{s. t.} \quad & \boldsymbol{q}_{i+1}^T = \boldsymbol{q}_i^T f_i(\hat{\boldsymbol{v}}, \boldsymbol{\Gamma}), \;\; i = 0, \ldots, M-1 \\
& \boldsymbol{q}_0 = \boldsymbol{q}_\mu \\
& \boldsymbol{g}_m(\boldsymbol{q}) + \mathbf{F}_m \boldsymbol{z}_m = \boldsymbol{y}_m \\
& \boldsymbol{q} \geq 0 \\
& \hat{\boldsymbol{v}} \geq 0,
\end{aligned} \tag{57}$$

where $h(\hat{\boldsymbol{v}})$ is the discretized version of

$$h(v) = \frac{1}{2}\int_0^{t_M}\int_{x_1}^{x_N} \left(a''(\tau,x)\left(\frac{\partial^2 v}{\partial \tau^2}\right)^2 + b''(\tau,x)\left(\frac{\partial^2 v}{\partial x^2}\right)^2 \right) d\tau dx. \tag{58}$$

$h(v)$ ensures that we only consider realistic surfaces, while the second component in the objective makes sure that the surface is consistent with observed market prices. The first constraint in (57) is the discretization of the Fokker-Planck equation, and is used to

generate prices of a collection of options from the current local volatility surface. The third constraint is used to ensure that the surface is consistent with observed market prices, while the last two constraints guarantee that the surface is free of strong arbitrage. The optimization problem (58) is a nonlinear and non-convex optimization problem. Even though the problem is non-convex, extensive testing has shown that we reach broadly the same solution which is stable for a time series of option data sets. Thus, the solution does not jump between very different local optima when the estimation problem is solved for many dates in a time series study, but there may exist other local optima. When implementing the model we have used AMPL together with the open source non-linear solver IPOPT (Wächter and Biegler, 2006).

4.2. **Recovering the real-world density.** When recovering the real-world PDF from observed option data we rely on a result derived by Heath and Platen (2006), which describes a direct relation between the real-world PDF and observed call option prices. For an index that pays a continuous deterministic dividend yield we have that the real-world PDF is given by the following expression

$$p_S(t, S_t, T, S_T) = \frac{S_T \mathrm{e}^{\int_t^T \delta_s ds}}{S_t} \left. \frac{\partial^2 C(t, S_t, T, K)}{\partial K^2} \right|_{K=S_T}, \tag{59}$$

where $C(t, S_t, T, K)$ is the price, at time t, of a call option with maturity T and strike price K. In equation (59), $p_S(t, S_t, T, S_T)$ denotes the real-world probability density of the index moving from the index level S_t at time t to S_T at time T. Since the expression for the RND as a function of the index value at time T is given by

$$q_S(t, S_t, T, S_T) = \mathrm{e}^{\int_t^T r_s ds} \left. \frac{\partial^2 C(t, S_t, T, K)}{\partial K^2} \right|_{K=S_T}, \tag{60}$$

we have the following connection between the real-world PDF and the formally obtained RND

$$p_S(t, S_t, T, S_T) = \mathrm{e}^{\int_t^T (\delta_s - r_s) ds} \frac{S_T}{S_t} q_S(t, S_t, T, S_T). \tag{61}$$

Hence, by using the RND estimated from observed option prices using the method in Section 4.1, we can directly recover also the real-world PDF by using the relationship in equation (61). As argued in Section 4.1, we have not restricted the model for the underlying index to be given by a local volatility function model. Thus, the volatility process for the index, which is assumed to approximate the numéraire portfolio, is not inconsistent with that of the generalized MMM. For the case when the underlying index is assumed to follow a local volatility function model, the consistency with the volatility process of the generalized MMM is supported by a result in Gyöngy (1986). By using this result, we can project the generalized MMM into a Markovian diffusion with a local volatility function, making the approach consistent also for this case.

The parameters of the generalized MMM are estimated with Simulated Maximum Likelihood, which requires the discretized real-world density instead of the continuous real-world density given in equation (59). The connection between the continuous RND and the discretized RND, $q_{i,j}$, at the point (T_i, S_j) on the grid is given by

$$q_S(t, S_t, T_i, S_j) = \frac{2}{S_{j+1} - S_{j-1}} q_{i,j}, \tag{62}$$

or equivalently

$$q_{i,j} = \frac{S_{j+1} - S_{j-1}}{2} q_S(t, S_t, T_i, S_j), \tag{63}$$

(see Andersen and Brotherton-Ratcliffe, 1997). Using the result for the continuous RND and real-world PDF we use the following transformation going from the discrete RND to the discrete real-world PDF

$$p_{i,j} = \frac{S_j}{S_t} \mathrm{e}^{\int_t^{T_i} (\delta_s - r_s) ds} q_{i,j}. \tag{64}$$

The discretized real-world PDF implied by observed option prices will then be used as input to the SML algorithm used for estimating the option implied MMM parameters.

4.3. **Simulated Maximum Likelihood.** Now that we have described how to estimate the discrete real-world PDF implied by observed option prices, we are ready to formulate the Simulated Maximum Likelihood (SML) problem that we set up to estimate the parameters of the MMM. Since we do not know the likelihood function in closed form we have to resort to a simulation based method in order to approximate the likelihood function. The method that we use is based on the method described in the recent paper by Kristensen and Shin (2012), where the authors use a non-parametric kernel based method to approximate the likelihood function. Since we do not have any actual observations of the index level as input to the MLE, but instead know the implied probabilities of the index levels on the grid at a specific future time, the method that we use is slightly altered compared to what is described in Kristensen and Shin (2012). The simulated values of the index that we use in the SML algorithm are generated with the method described in Rendek (2013).

In our case we have the estimated option implied probabilities of observing outcomes for the index at a specific future time, T_i, for different values on the grid given by the pairs

$$\{S_j, p_{i,j}\}_{j=1}^N. \tag{65}$$

Since the MMM models the total return index level, $\tilde{S}_t$, we will first transform these pairs onto the grid containing the values for the total return index

$$\left\{\tilde{S}_j, \tilde{p}_{i,j}\right\}_{j=1}^N. \tag{66}$$

Introducing the function $g(T, S_T)$, such that $\tilde{S}_T = g(T, S_T) = \mathrm{e}^{\int_0^T \delta_s ds} S_T$, we have that

$$\tilde{p}_{i,j} = \mathbb{P}(\tilde{S}_{T_i} = \tilde{S}_j) = \mathbb{P}(g(T_i, S_{T_i}) = \tilde{S}_j) = \mathbb{P}(S_{T_i} = g^{-1}(T_i, \tilde{S}_j) = \mathbb{P}(S_{T_i} = S_j) = p_{i,j}, \tag{67}$$

and hence the pairs in equation (66) are equivalent to the pairs

$$\left\{\tilde{S}_j, p_{i,j}\right\}_{j=1}^N. \tag{68}$$

Now, assume that we have N_{obs} observations of the underlying total return index level, where the observations are distributed among the grid points $\{\tilde{S}_j\}_{j=1}^N$. We let $n_{i,j}$, $j = 1, \ldots, N$, denote the number of observations that assume the value $\tilde{S}_j$ at time T_i. Thus, the likelihood function given the outcomes for the total return index level is given by

$$L(\boldsymbol{\theta}) = \prod_{j=1}^N f(\tilde{S}_j; \boldsymbol{\theta})^{n_{i,j}}, \tag{69}$$

where $f(\cdot;\boldsymbol{\theta})$ is the continuous PDF as a function of the parameter values $\boldsymbol{\theta}$. Consequently, the log-likelihood function is given by

$$l(\boldsymbol{\theta}) = \ln(L(\boldsymbol{\theta})) = \sum_{j=1}^{N} n_{i,j} \ln(f(\tilde{S}_j;\boldsymbol{\theta})). \tag{70}$$

The maximum likelihood estimate of $\boldsymbol{\theta}$, scaled by $\frac{1}{N_{\text{obs}}}$, is now given by

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}} \frac{1}{N_{\text{obs}}} \sum_{j=1}^{N} n_{i,j} \ln(f(\tilde{S}_j;\boldsymbol{\theta})), \tag{71}$$

which for the limit $N_{\text{obs}} \to \infty$ gives

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}} \sum_{j=1}^{N} p_{i,j} \ln(f(\tilde{S}_j;\boldsymbol{\theta})), \tag{72}$$

where

$$\boldsymbol{\Theta} = \{a \geq 1;\, \epsilon > 0;\, \gamma > 0;\, A \geq 0;\, M_0 > 0;\, Y_0 > 0\}. \tag{73}$$

Due to the restricted data set available, the parameters a, A and Y_0 are not safely steered towards plausible values by the optimization, so we will instead estimate these parameters from historical index data using the method described in Platen and Rendek (2012a). These parameter values are then treated as constants in the optimization. The reduced parameter set that we solve for is thus given by

$$\boldsymbol{\Theta} = \{\epsilon > 0;\, \gamma > 0;\, M_0 > 0\}. \tag{74}$$

In Section 2.3 we argued that the properties of the optimization problem are improved if we redefine the parameter set according to

$$\boldsymbol{\Theta} = \left\{\epsilon > 0;\, \frac{\epsilon}{\gamma} > 0;\, M_0 > 0\right\}, \tag{75}$$

and this is the actual parameter set that we solve for when we use SML to estimate the option implied MMM parameters.

Since we do not have the PDF as a function of the parameter values in closed form, we need to have a method to estimate the PDF from simulated index values. As in Kristensen and Shin (2012) we will use a non-parametric kernel based method to numerically approximate the PDF. Assume that we have simulated n_{sim} i.i.d. outcomes for the total return index $\{\tilde{S}_k^{\theta}\}_{k=1}^{n_{\text{sim}}}$. By construction it follows that the n_{sim} simulated i.i.d. random variables $\{\tilde{S}_k^{\theta}\}_{k=1}^{n_{\text{sim}}}$ follow the target distribution

$$\tilde{S}_k^{\theta} \sim f(\cdot;\boldsymbol{\theta}). \tag{76}$$

They can, therefore, be used in order to estimate $f(\tilde{S};\boldsymbol{\theta})$ with the help of kernel based methods. Define the kernel density estimator

$$\hat{f}(\tilde{S};\boldsymbol{\theta}) = \frac{1}{n_{\text{sim}}} \sum_{k=1}^{n_{\text{sim}}} K_h(\tilde{S} - \tilde{S}_k^{\theta}), \tag{77}$$

where

$$K_h(x) = \frac{K(x/h)}{h}, \tag{78}$$

and where $K(\cdot)$ is the kernel function, $K_h(\cdot)$ is the scaled kernel function and $h > 0$ is the bandwidth. When we implement the kernel estimator for the estimation we will use a Gaussian kernel. Thus we have that the scaled kernel function is given by

$$K_h(\tilde{S} - \tilde{S}_k^\theta) = \frac{1}{\sqrt{2\pi}h} \exp\left(\frac{-(\tilde{S} - \tilde{S}_k^\theta)^2}{2h^2}\right), \tag{79}$$

and hence the kernel density estimator is given by

$$\hat{f}(\tilde{S}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}h n_{\text{sim}}} \sum_{k=1}^{n_{\text{sim}}} \exp\left(\frac{-(\tilde{S} - \tilde{S}_k^\theta)^2}{2h^2}\right). \tag{80}$$

It can be shown that if Gaussian kernel functions are used to approximate univariate data and when the underlying density being estimated is Gaussian, then the optimal choice for the bandwidth is given by

$$h = \left(\frac{4\hat{\sigma}^5}{3n_{\text{sim}}}\right)^{\frac{1}{5}} \approx 1.06\,\hat{\sigma}(n_{\text{sim}})^{-\frac{1}{5}}, \tag{81}$$

where $\hat{\sigma}$ is the sample standard deviation of the simulated values of $\tilde{S}$ (see e.g. Silverman, 1998). When we implement the kernel density estimator in the optimization we will use this value for the bandwidth even though the underlying density is not Gaussian. This will, thus, introduce a bias in the estimation and there is hence room for improvements in the estimation when choosing the optimal bandwidth.

After equation (80) has been used to obtain the simulated density, we can construct the following simulated maximum likelihood estimate for the true value of the parameters, $\boldsymbol{\theta}_0$,

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}}\ \hat{l}(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}} \sum_{j=1}^{N} p_{i,j} \ln(\hat{f}(\tilde{S}_j; \boldsymbol{\theta})). \tag{82}$$

When we solve the optimization problem (82) numerically we use the same generated random variables for all values of $\boldsymbol{\theta}$, i.e. for every iteration in the optimization. From equation (77) we have that if K_h and $\tilde{S}_k^\theta$ are continuously differentiable in $\boldsymbol{\theta}$, then $\hat{l}(\boldsymbol{\theta})$ has the same property. This follows from the chain rule and the fact that we use the same random draws for all values of $\boldsymbol{\theta}$. Since in our case both K_h and $\tilde{S}_k^\theta$ are continuously differentiable in $\boldsymbol{\theta}$ then this property also holds for $\hat{l}(\boldsymbol{\theta})$.

4.3.1. *Altering the objective function for the SML estimation.* In the previous subsection we have described the optimization problem that we solve in the SML estimation of the model parameters. The description was based on that we fit the model implied PDF to the market implied PDF for one specific maturity. Including PDFs for more than one maturity in the objective function for the SML based optimization, would let us also consider the dynamics for the PDFs that are implied by the observed prices of liquid options. If we let n_τ denote the number of maturities included in the objective function,

and let $i \in \{1, \ldots, n_\tau\}$ denote the indices of the included maturities, we can formulate the following alternative optimization problem for the estimation of the model parameters

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}}\ \hat{l}(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}} \sum_{i=1}^{n_\tau} w_i \left(\sum_{j=1}^{N} p_{i,j} \ln(\hat{f}_i(\tilde{S}_j; \boldsymbol{\theta})) \right), \tag{83}$$

where we will typically let the weights w_i be given by $1/n_\tau$. In equation (83) we have that $\hat{f}_i$ denotes the kernel density estimator of the model implied PDF for the ith maturity.

5. Conclusions

In this paper we have developed a methodology for simultaneous estimation of the real-world probability density and liquidity premia from observed index option data. The methodology is based on the possibility to recover the real-world density from liquid options, i.e. options with strikes that are not too far from the index level and that have shorter term maturities. In order to recover the real-world density we use a recently developed method for non-parametric estimation of the local volatility surface, where the risk-neutral density is a by-product of the estimation. It is vital that the local volatility estimation extracts a plausible and realistic risk-neutral density, and is able to separate out the noise that is contained in option price data. The estimation method that we have used has proven to be able to produce risk-neutral densities which are consistent with market prices while also being realistic and stable over time. Given the high quality estimates of the risk-neutral density we have then used the results from Heath and Platen (2006) in order to directly recover the market implied real-world density. The local volatility estimation framework is presented as an advanced interpolation tool for producing smooth and plausible call price surfaces. The recovery of the real-world densities does hence not rely on assuming a local volatility function model for the underlying index. From recovered real-world densities we have then shown how to estimate the option implied parameters of the generalized Minimal Market Model. Assuming that this model provides an accurate description of the market dynamics of the underlying index, the option implied parameters do not only reflect forward looking information, but are also able to connect to the historically observed characteristics of the index. These properties of the model have allowed us to both extract liquidity premia embedded in option prices, and provide a more realistic description of the real-world density.

References

Andersen, L.B.G. and Brotherton-Ratcliffe, R., The equity option volatility smile: an implicit finite-difference approach. *Journal of Computational Finance*, 1997, **1**, 5–37.

Ané, T. and Geman, H., Order flow, transaction clock and normality of asset returns. *Journal of Finance*, 2000, **60**, 2259–2284.

Baldeaux, J., Grasselli, M. and Platen, E., Pricing currency derivatives under the benchmark approach. *Journal of Banking and Finance*, 2015, **53**, 34–48.

Barkhagen, M. and Blomvall, J., Non-parametric estimation of stable local variance surfaces. Working paper, Linköping University, 2015.

Björk, T., *Arbitrage Theory in Continuous Time* 3^d *Ed.*, 2011 (Oxford University Press: New York).

Bollen, N.P.B. and Whaley, R.E., Does net buying pressure affect the shape of implied volatility functions?. *Journal of Finance*, 2004, **59**, 711–753.

Bouchouev, I. and Isakov, V., Uniqueness, stability and numerical methods for the inverse problem that arises in financial markets. *Inverse problems*, 1999, **15**, 95–116.

Breeden, D.T. and Litzenberger, R.H., Prices of state-contingent claims implicit in option prices. *Journal of Business*, 1978, **51**, 621–651.

Carr, P. and Yu, J., Risk, return, and Ross recovery. *Journal of Derivatives*, 2012, **20**, 38–59.

Coleman, T.F., Li, Y. and Verma, A., Reconstructing the unknown local volatility function. *Journal of Computational Finance*, 1999, **2**, 77–102.

Derman, E. and Kani, I., Stochastic implied trees: Arbitrage pricing with stochastic term and strike structure. *International Journal of Theoretical and Applied Finance*, 1998, **1**, 61–110.

Gârleanu, N., Pedersen, L.H. and Poteshman, A.M., Demand-based option pricing. *Review of Financial Studies*, 2009, **22**, 4259–4299.

Gyöngy, I., Mimicking the one-dimensional marginal distributions of processes having an Ito differential. *Probability Theory and Related Fields*, 1986, **71**, 501–516.

Heath, D. and Platen, E., Local volatility function models under a benchmark approach. *Quantitative Finance*, 2006, **6**, 197–206.

Jarque, C.M. and Bera, A.K., Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 1981, **7**, 313–318.

Kristensen, D. and Shin, Y., Estimation of dynamic models with nonparametric simulated maximum likelihood. *Journal of Econometrics*, 2012, **167**, 76–94.

Lagnado, R. and Osher, S., A technique for calibrating derivative security pricing models: numerical solution of the inverse problem. *Journal of Computational Finance*, 1997, **1**, 13–25.

Lilliefors, H., On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 1967, **62**, 399–402.

Loewenstein, M. and Willard, G.A., Local martingales, arbitrage, and viability: free snacks and cheap thrills. *Economic Theory*, 2000, **16**, 135–161.

Long, J.B., The numeraire portfolio. *Journal of Financial Economics*, 1990, **26**, 29–69.

Platen, E., Arbitrage in continuous complete markets. *Advances in Applied Probability*, 2002, **34**, 540–558.

Platen, E., Diversified portfolios with jumps in a benchmark framework. *Asia-Pacific Financial Markets*, 2005, **11**, 1–22.

Platen, E., Real world pricing of long term constracts. *UTS Research Paper Series*, 2009.

Platen, E., A benchmark approach to investing and pricing. In *The Kelly capital growth investment criterion*, edited by L. MacLean, E.O. Thorp and W. Ziemba, pp. 409–425, 2011, World Scientific.

Platen, E. and Heath, D., *A Benchmark Approach to Quantitative Finance*, 2006, Springer: Berlin.

Platen, E. and Rendek, R., The affine nature of aggregate wealth dynamics. *UTS Research Paper Series*, 2012a.

Platen, E. and Rendek, R., Approximating the numéraire portfolio by naive diversification. *Journal of Asset Management*, 2012b, **13**, 34–50.

Rendek, R., Modeling diversified equity indices. PhD thesis, University of Technology, Sydney, 2013.

Ross, S., The recovery theorem. *Journal of Finance*, 2015, **70**, 615–648.

Silverman, B.W., *Density Estimation for Statistics and Data Analysis*, 1998 (Chapman and Hall/CRC: Boca Raton, FL).

Stephens, M.A., EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 1974, **69**, 730–737.

Wächter, A. and Biegler, L.T., On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 2006, **106**, 25–57.

Waldén, J., Recovery with unbounded diffusion processes. Working paper, Haas School of Business, University of California at Berkeley, 2013.

Appendix A. Data

We will here describe the option data set that was used for the illustration of the estimated option implied surfaces that was presented in Section 2.1. The options are cash settled European style options traded on the Chicago Board Options Exchange (CBOE). The option data was obtained from Thomson Reuters Datastream together with the corresponding closing values for the S&P 500 index. On 3 October 2014 option series with the following 3 maturity dates were listed on CBOE, and had associated bid-ask quotes for options belonging to the subset $\mathcal{L}$ (described in Section 2.1): 16 October 2014, 20 November 2014, 18 December 2014. The option data belonging to the set $\mathcal{L}$ for this date consists of 451 contracts divided between 232 calls and 219 puts. S&P 500 futures are traded on the Chicago Mercantile Exchange (CME), and the closing prices of these futures are not synchronized with the closing prices of the CBOE index options. Thus, we do not use bid and ask quotes of traded futures as input to the estimation. Instead we use the futures prices implied by put-call parity for the 5 strikes closest to the current index level as input. The trading hours for S&P 500 index options on the CBOE are between 8:30 a.m. and 15:15 p.m. Chicago time. The constituents of S&P 500 are traded on the New York Stock Exchange (NYSE) and the NASDAQ Stock Market (NASDAQ). The trading hours for the stocks on the NYSE are between 9:30 a.m. and 16:00 p.m. New York time. Since the time difference between New York and Chicago is one hour, the closing index value of S&P 500 is recorded 15 minutes before the end of trading for the options on the CBOE. This means that we will introduce a small error in the estimated implied RND since the starting value for the RND in the estimation will be set to one for the closing index value for each date. This effect is however limited and will not affect the mean of the estimated RND from the first option maturity and forward on the grid, since we are using futures prices implied by put-call parity as input.

The interest rates are zero yields collected from Thomson Reuters Eikon, with time-to-maturity from 1 day up to 6 months. The continuously compounded interest rates corresponding to the option maturities are then obtained by a cubic spline interpolation.

Mathias Barkhagen and Jörgen Blomvall, Division of Production Economics, Department of Management and Engineering, Linköping University, 581 83 Linköping, Sweden
E-mail address: mathias.barkhagen@liu.se

E-mail address: jorgen.blomvall@liu.se

Eckhard Platen, University of Technology, Sydney, Finance Discipline Group and School of Mathematical Sciences, PO Box 123, Broadway, NSW 2007, Australia
E-mail address: eckhard.platen@uts.edu.au