

Estimating option-implied distributions in illiquid markets and implementing the Ross recovery theorem

By Emlyn Flint and Eben Maré

Submission date 1 March 2017

Acceptance date 13 September 2017

ABSTRACT

In this research we describe how forward-looking information on the statistical properties of an asset can be extracted directly from options market data and demonstrate how this can be practically applied to portfolio management. Although the extraction of a forward-looking risk-neutral distribution is well-established in the literature, the issue of estimating distributions in an illiquid market is not. We use the deterministic SVI volatility model to estimate weekly risk-neutral distribution surfaces. The issue of calibration with sparse and noisy data is considered at length and a simple but robust fitting algorithm is proposed. We further attempt to extract real-world implied information by implementing the recovery theorem introduced by Ross (2015). Recovery is an ill-posed problem that requires careful consideration. We describe a regularisation methodology for extracting real-world implied distributions and implement this method on a history of SVI volatility surfaces. We analyse the first four moments from the implied risk-neutral and real-world implied distributions and use them as signals within a simple tactical asset allocation framework, finding promising results.

KEYWORDS

Option-implied distributions; SVI volatility model; Ross recovery theorem; Tikhonov regularisation; illiquid derivative markets

CONTACT DETAILS

Mr Emlyn Flint, Peregrine Securities, The Citadel, 15 Cavendish Street, Cape Town

Email: emlynf@peregrine.co.za; Department of Actuarial Science, University of Pretoria, Private Bag X20, Pretoria

Prof Eben Maré, Department of Mathematics and Applied Mathematics, University of Pretoria, Private Bag X20, Pretoria; Email: eben.mare@up.ac.za

1. INTRODUCTION

1.1 An important requirement for optimal portfolio construction is an understanding of the future possible returns of the constituent assets. Armed with this understanding, portfolio managers can ensure that their chosen combination of assets will lead to a portfolio that is consistent with investors' risk tolerances and return objectives. Unfortunately, forecasting return distributions accurately is a challenging endeavour. A common approach is to use historical data as the basis for forecasts. For example, standard deviation and expected returns are easily estimated from historical data and, when combined with an assumption of normally distributed returns, provide a completely specified return distribution. Unfortunately, empirical studies have shown that expected return and standard deviation estimated from historical data are unstable and the assumption that historical estimates will apply at a future date corresponding to the investment horizon is questionable at best (see DeMiguel et al. (2009) and the references therein). An alternate forecasting method is to extract forward-looking information on the statistical properties of an asset directly from options market data related to that asset.

1.2 The seminal work of Black & Scholes (1973) and Merton (1973) proved that the value of an option in a complete market was independent of the expected return on the underlying asset and thus gave rise to the risk-neutral valuation framework. Under this framework, the only unknown parameter affecting an option's value is the volatility of the underlying asset. Because of this, the Black–Scholes–Merton (BSM) pricing formula has become ubiquitous in derivatives markets worldwide due to its ability to monotonically translate any option price into a single, easily-comparable, implied volatility value. In this sense, implied volatility is the one language common to all option markets.

1.3 Implied volatility surfaces vary, in practice, in three important ways from the flat theoretical BSM surface. Firstly, implied volatility varies with the strike of an option. Secondly, implied volatility varies with option term. Thirdly, the shape of the volatility surface changes over time depending on the underlying market regime and trading conditions. There are also certain practical violations of the BSM theory—e.g. non-zero transactions, discrete trading time limits, and finitely divisible prices—which all generally have the effect of raising the empirical implied volatility surface to some extent. It is generally accepted that the volatility surface represents a combination of the consensus view of the terminal asset return distribution, current market risk preferences, and any supply-demand factors stemming from structural market issues (i.e., option liquidity premia). Therefore, in addition to providing one with a means of trading and pricing options, the implied volatility surface can be viewed as containing the sum of all forward-looking information known (or assumed) about the underlying asset.

1.4 The idea of accessing this embedded information is not new. Implied volatility has long been used as a broad gauge of investor risk sentiment or fear, with the Chicago Board

of Options Exchange Volatility Index (VIX) being the most commonly referenced measure today. However, since the mid 1990s, options have increasingly become assets in their own right and there has been a concerted effort to study the extent of the predictive information embedded in these markets.

1.5 Breeden & Litzenberger (1978) proved that the forward-looking risk-neutral return distribution (RND) could be extracted directly from an arbitrage-free derivatives market provided that one knows the European option prices for all levels of the underlying. This result gave investors the means to estimate the forward-looking RND for a given term from the implied volatility skew for that same term. A range of interesting statistical metrics can then be calculated from the RND, including volatility (e.g. the VIX), skewness, kurtosis, and value-at-risk, all of which are inherently forward-looking by construction. Furthermore, if volatility skews on an equity index and its underlying stocks are available, then it is also possible to estimate forward-looking implied stock correlations and betas. Kempf et al. (2015), Baule et al. (2015), DeMiguel et al. (2013), Buss & Vilkov (2012) and Kostakis et al. (2011), *inter alia*, have shown that such implied moments and statistics significantly outperform their historical counterparts in a range of portfolio, risk management and trading applications.

1.6 An important point to remember is that risk-neutral probabilities are not equivalent to real-world probabilities. A change in risk-neutral probabilities can either stem from a change in the underlying real-world probabilities or from a change in underlying risk preferences (Malz, 2014). Furthermore, RNDs do not give one any forward-looking information on the real-world expected return. In fact, until very recently it was considered impossible to extract any real-world information directly from option markets without either having to make stringent assumptions about investors' risk preferences or resorting to estimation of the same from historical data. However, a recent development has brought this belief into contention.

1.7 Ross (2015) postulated the recovery theorem, which, for a given set of market and risk preference restrictions, makes it possible to estimate real-world information directly from an options market. Although some of the assumptions underlying Ross's recovery theorem are obvious simplifications of actual market conditions, the more pertinent practical question—as Audrino et al. (2015) point out—is whether this recovered real-world distribution provides any additional insight beyond that found in its more easily available risk-neutral counterpart. While a number of researchers have raised fundamental questions as to whether there can actually be either unique or better information within a secondary market, the fact of the matter is that option-implied information is currently used in a wide range of market applications. Therefore, robust estimation of this information is a critical empirical issue. In this work, we contribute to the research in this field by considering, in detail, the estimation and application of risk-neutral and real-world option-implied distributions in an illiquid market setting where data are both sparse and noisy.

1.8 The remainder of this paper is set out as follows. Section 2 tackles the problem of estimating RNDs by introducing a range of common estimation techniques and discussing their applicability in the context of the illiquid South African options market. The optimal technique as selected by the authors is then discussed at length and a practical RND estimation algorithm is presented. Section 3 introduces the recovery theorem, discusses some implementation challenges and presents an algorithm for applying the recovery theorem using regularised least squares. Empirical results using South African options data are presented in Section 4. General option-implied data applications are discussed, recovered real-world distribution moments are compared to their risk-neutral counterparts and a tactical asset allocation example using implied information is presented. Section 5 highlights conclusions reached by the authors.

2. ESTIMATING OPTION-IMPLIED DISTRIBUTIONS IN ILLIQUID MARKETS

2.1 In a complete and arbitrage-free market, Cox & Ross (1976) show that the model-free value of a European call option C_t at time t , with expiry T , term $\tau = T - t$ and strike K is equal to

$$C_t(K, T) = e^{-r\tau} \int_0^{\infty} (S_T - K)^+ q(S_T) ds, \quad (1)$$

where r is the risk-free rate, S_T is the terminal price of the underlying and $q(S_T)$ is the terminal risk-neutral distribution of the underlying. Taking the second derivative with respect to strike yields the seminal result from Breeden & Litzenberger (1978):

$$e^{r\tau} \frac{\partial^2 C}{\partial K^2} = q(K). \quad (2)$$

2.2 In theory, one needs a continuum of option prices across strike levels for a given term in order to calculate the RND. In practice though, only a discrete set of options are actually traded and thus some estimation procedure is required. A wide range of RND estimation techniques have been suggested in the literature, which can be broadly classified by whether they work with equations 1 or 2.¹

2.3 The techniques based on equation 1 postulate some distributional form for the RND, which is then evaluated based on an objective function measuring the distance between the estimated and actual option prices. Parametric forms include adding expansion terms to base distributions (Coutant et al., 1998), using complex underlying distributions (Posner & Milevsky, 1998), or using mixtures of lognormal distributions (Melick & Thomas, 1997).

¹ Obviously, there are some techniques that cannot easily be shoehorned into this classification scheme but we believe that it still provides a useful means of summarising the most popular techniques.

Nonparametric forms include the use of entropy-based methods (Buchen & Kelly, 1996) and kernel-based methods (Aït-Sahalia & Lo, 1998).

2.4 The estimation techniques based on equation 2 instead postulate some continuous form of the underlying volatility skew which can be used to interpolate between the traded option volatilities—and thus prices—as well as extrapolate outside of the traded strike range. The second derivative of the call prices and hence the RND is then found numerically. These techniques can be divided into three sub-categories. Firstly, one can use curve-fitting techniques such as cubic splines to interpolate between and extrapolate from traded option volatilities (Bliss & Panigirtzoglou, 2002). Secondly, one can fit deterministic volatility models to the traded data (Shimko, 1993; Dumas et al., 1998), and thirdly, one can postulate more complex models for the underlying process in the form of stochastic volatility (Heston, 1993), jump diffusion (Merton, 1976), or a combination of the two (Bates, 2000).

2.5 Given this vast array of RND estimation techniques as well as the range of different applications in which these RNDs are used, it is perhaps not surprising that it remains an open question as to which technique—if any—is considered ‘optimal’. From the small number of comparative studies undertaken to date, the only universal conclusion is that estimating RNDs is an ill-posed problem, which can be highly dependent on the estimation technique as well as the available data.² This means that the choice of technique should be considered an active decision in the RND estimation process as well as in the larger recovery process.

2.6 These difficulties relating to estimating RNDs become even more important in illiquid markets where option data are both sparse and noisy. Because of these two features, many of the techniques proposed above become unsuitable. Apart from the shape-constrained kernel method of Aït-Sahalia & Duarte (2003), the majority of the distributional-based methods are largely unconstrained and thus will struggle under sparse, noisy data conditions as their inherent flexibility may actually augment estimation error. For example, Cooper (1999) shows that under real-world conditions, noisy option data can lead to large spikes in the RNDs estimated using the ‘mixture of lognormals’ approach. McManus (1999) notes a similar result for entropy-based techniques. The same argument can be extended to spline-based techniques, which are heavily dependent on the choice and number of knots. Spline-based techniques also raise the additional question of how to extrapolate implied volatility skews beyond the traded range. Malz (2014) and Bliss and Panigirtzoglou (2002; 2004) suggest assuming flat volatility—and thus lognormal RND tails—outside of the traded range, whereas Figlewski (2008) instead suggests grafting Generalised Extreme Value distribution tails onto the estimated central portion of the RND. In either case, the researcher is ultimately pre-specifying the tail structure of the RND, thereby actually making the spline technique somewhat parametric.

2 See, for example, Aït-Sahalia & Duarte (2003), Cooper (1999), McManus (1999) and Coutant et al. (1998).

2.7 There is also another issue that needs to be considered in this research. Successful application of the recovery theorem requires RNDs across a number of terms—i.e. an RND ‘surface’—rather than the single-term distributions usually considered in the literature. This means that not only does one have to consider arbitrage constraints across strike but also across term. While the kernel methods of Aït-Sahalia & Duarte (2003) do consider the former, they do not formally make provision for the latter. In contrast, the problem of static arbitrage across both strike and term has been extensively researched in the volatility modelling literature.³ Therefore, given our requirement of a complete arbitrage-free RND surface, this would suggest choosing either the deterministic or stochastic volatility modelling approach. Popular candidates in each area are Gatheral’s (2004) stochastic volatility inspired (SVI) model and Heston’s (1993) stochastic volatility model respectively.

2.8 Gatheral (2011) states that stochastic volatility models fail to capture the dynamics of short-term volatility skews and can also be hard to calibrate in practice. On the other hand, the comprehensive study by Tompkins (2001) suggests that most option markets are well modelled by simpler deterministic functions. Furthermore, deterministic models provide not only the flexibility to calibrate implied volatility separately across strike and term but also the simplicity to ensure arbitrage-free volatility surfaces with minimal model error. Based on these observations, as well as our overarching aim to extract information in as robust and flexible a manner as possible, we choose to model the implied volatility surface—and thus the RND surface—using the SVI model.

2.9 The South African Options Market

2.9.1 In this study, we consider fully margined options on Top40 index futures, traded on the South African Futures Exchange (SAFEX).⁴ These listed options expire quarterly on the third Thursday of March, June, September, and December each year. West (2005) provides one of the few studies on the volatility calibration challenges faced within this market. At the time of his study, over the counter (OTC) structures comprised the majority of South African (SA) option market activity. However, the subprime crisis in 2008 changed the manner in which investors traded. A renewed interest in regulation together with a greater appreciation of credit risk resulted in a significant increase in the number of exchange-traded contracts and a concomitant decline in OTC volumes. This trend was further aided by the introduction of the SAFEX Can-Do derivatives platform, which essentially gave investors the ability to list, trade and margin any exotic derivative directly with SAFEX. The shift towards exchange-traded contracts also meant a far larger proportion of the underlying trade data became publicly available.

2.9.2 In the SA market, participants mostly use derivatives as hedging instruments, meaning that open interest is concentrated in put options with strikes below current index

3 For example, see Carr et al. (2003), Carr & Madan (2005) and Roper (2010).

4 The Top40 currently comprises the largest 43 South African stocks listed by gross market capitalisation.

levels. These hedging structures are usually short-term with volumes concentrated in the three closest expiries—i.e. up to nine months. Trade in any instruments with term beyond 15 months is extremely rare. The size of such hedge trades can also, at times, dominate all other trades, leading to an extremely skewed open interest distribution across option strikes. That being said, total option volumes traded in SA remain small in comparison to other developed markets. On any given day, the number of trades varies significantly and there could even be no trades across any expiry. The traded strike range is also quite narrow and generally spans a maximum range of -20% to $+15\%$ of current (spot) index levels.

2.9.3 Daily listed Top40 option trade data is freely available from the Johannesburg Stock Exchange (JSE) website dating back to February 2011. We further sourced option trade data back to September 2005 from Peregrine Securities, a large derivatives broker in South Africa. For each option trade, the full dataset generally includes trade date and time, futures level, strike, traded volatility, price, option type, and volume.⁵ Market participants do have some leeway in terms of what and how to report this information to the exchange and so incomplete records can and do occur.

2.10 The Stochastic Volatility Inspired Model

2.10.1 The SVI model was disseminated by Gatheral (2004) and subsequently has arguably become the practitioner's model of choice in the equity derivative space. It is known to fit equity volatility skews extremely well but is still intuitive and easy to implement. Denoting the futures level as F , the term as τ and the strike as K , we can write the SVI implied variance as

$$\sigma^2(x, \tau) = a + b \left(\rho(x - m) + \sqrt{(x - m)^2 + s^2} \right), \quad (3)$$

where $x = \ln\left(\frac{K}{F}\right)$ is the log-moneyness and the parameter set $\{a, b, \rho, m, s\}$ is specific to each expiry. This parameterisation was inspired by the large-term asymptotic behaviour of the Heston stochastic volatility model. In essence, the SVI model fits a hyperbola to implied variance in log-moneyness space. This particular form is chosen because it ensures that variance is linear as $|x| \rightarrow \infty$ (a fundamental characteristic of volatility surfaces) while still being convex around the at-the-money (ATM) level. This is intuitive for traders in that the more out-the-money (OTM) an option is, the more volatility convexity the option displays. Each SVI parameter has an intuitive geometric interpretation (Gatheral & Jacquier, 2014):

- a defines the overall level of variance and shifts the skew vertically;
- b defines the angle between the left- and right-side variance slopes;
- ρ rotates the variance curve clockwise around the current forward level;
- m shifts the variance curve left or right;
- s defines the amount of ATM variance curvature.

5 Note that there is no bid-ask spread information included.

2.10.2 This particular choice of parameterisation coupled with the five degrees of freedom generally ensures an extremely good fit in practice, particularly in the equity index space. Furthermore, because of its characterisation, the SVI model is able to provide decent approximations for deep OTM volatilities and can also produce sufficient ATM curvature at very short terms, a known failing of many stochastic volatility models. Finally, Gatheral & Jacquier (2014) also show that the SVI single-expiry calibration process is also easily coupled with calendar-spread arbitrage checks, which enables straightforward construction of smooth, arbitrage-free implied volatility surfaces.

2.10.3 One drawback of the SVI model is that the usual least squares minimisation of the implied volatility objective function is very sensitive to initial parameter guesses. Furthermore, the function displays several local minima, which can seriously bias final parameter estimates. DeMarco & Martini (2009) addressed this shortcoming by finding a robust quasi-explicit calibration process which produced a reliable and stable parameter set. Through a clever change of variables, the initial five-dimensional SVI minimisation problem is recast into a much simpler two-dimensional problem, with the remaining three variables having quasi-explicit solutions within the new framework. This ingenious ‘2+3’ procedure is much less sensitive to initial guesses and provides stable, arbitrage-free SVI parameters. See Appendix A for more detail.

2.11 Constructing SVI Volatility and RND Surfaces

2.11.1 Although international literature on modelling implied volatility is vast, the majority of these studies are not easily applicable to the SA derivatives market due to its illiquid nature and fairly unique trading dynamics, as described above. To the authors’ knowledge, only West (2005) and Kotzé & Joseph (2009) discuss the issue of calibration in such a market. West (2005) calibrates the SABR model of Hagan et al. (2002) to options on Top40 index futures, while Kotzé & Joseph (2009) do the same but using a quadratic deterministic volatility model.⁶ Both studies stress the need for robust and sensible calibration algorithms and put forth several useful suggestions, which are incorporated below, for reaching that goal. Creating a robust calibration procedure still requires some “creative decision making” according to West.⁷ In this context, the SVI volatility and RND surface algorithm provided in this work represents a blend of theoretical best practices and market experience in the presence of severe practical constraints. For a given point in time, we construct implied volatility and RND surfaces as follows:

- Collate Top40 option trade data for the past seven days. Backfill missing values as required using the given information and the Black (1976) pricing equation adjusted for fully margined options. Discard those records which cannot be completed.
- Apply a daily exponential time-weighting function $\lambda=0.915$ to moderately down-weight older trades and a stepped size-weighting function to significantly up- or down-weight

6 This is the model currently used by SAFEX for mark-to-market and margining purposes.

7 Note that the term ‘robust’ is used throughout the paper purely in a model calibration context.

- trades falling in pre-specified size buckets ($w_i = \{0.1, 0.8, 1, 0.8\}$ for trades of less than 100, 500, 2 000 and 10 000 contracts respectively).⁸
- In cases of extreme data sparsity, include several OTM skew markers from the previous period's calibration, adjusted for the current ATM volatility level.
 - Calibrate the SVI model separately to each traded expiry using the '2+3' algorithm of DeMarco & Martini (2009).
 - Check for calendar-spread arbitrage by examining the total variance plot for any crossed lines. If necessary, recalibrate the SVI parameters from shortest to longest expiration and include a large penalty for crossing with the previous skew, as per Gatheral & Jacquier (2014).
 - Use the modified SVI parameters to create volatility skews across a 20–300% range of the prevailing forward prices.
 - Interpolate linearly in total variance space between the calibrated expiries to create monthly volatility skews ranging from 1–15 months (total range dependent on available expiries).
 - Calculate call prices across the full strike range at each term from the interpolated volatility skews and estimate the monthly RNDs numerically using equation 2.

2.11.2 We use this fitting procedure to create weekly arbitrage-free implied volatility and RND surfaces over the period 5 September 2005 to 16 May 2016, giving a total of 559 surface observations. The prevailing interest rate and dividend yield curves are also recorded at the calibration dates.

2.11.3 It is important to note that the only part of the recovery process directly susceptible to any data issues is the RND estimation. Any effects thereof will indirectly impact the recovery process via the estimated RND surface. Therefore, in order to create a recovery process which is robust to small samples, outliers and noisy data, it is necessary to create a robust RND estimation procedure. We take this opportunity to reiterate that the choice of RND estimation technique is an active decision in the larger recovery process.

3. RECOVERING REAL-WORLD IMPLIED DISTRIBUTIONS

We provide a brief outline below of the Ross recovery theorem along with its underlying assumptions in a style similar to Spears (2013).⁹ We then consider some of the technical difficulties in applying the recovery theorem in practice and present our implementation procedure. Note that the illiquid market data issues highlighted above do not directly affect the recovery process as the only required input is an estimated RND surface.

8 The choice of time-weight exponent is based on the calibration work of Kotze & Joseph (2009). In terms of the size weighting buckets, while the lower weighting for smaller trades is intuitive, a lower weight is also applied to ultra-large trades because such trades would generally have fewer potential market makers and thus pricing efficiency may decrease.

9 Interested readers can find further mathematical detail in Ross (2015). Extensions to the original theory are presented in Carr & Yu (2012), Dubynskiy & Goldstein (2013), Martin & Ross (2013), Walden (2014) and Borovička et al. (2016).

3.1 The Ross Recovery Theorem

3.1.1 Before stating the recovery theorem from Ross (2015), we need to introduce several underlying financial concepts. Assume that the underlying asset can only take on a finite n number of states. The transition probability matrix $P = (p_{ij})$ then defines how likely the underlying is to move from state i to another state j over the next time period. Assuming that these transition probabilities are time-homogeneous, we can write this mathematically as

$$p_{ij} = Pr(S_{t+1} = j | S_t = i) > 0 \quad \forall i, j \leq n, t > 0. \quad (4)$$

Note that if it is possible to reach any state from any other starting state given sufficient time, then P is said to be irreducible and it must therefore hold that $p_{ij}^t > 0$ for some t .

3.1.2 In this work, we will let P represent the transition probability matrix (TPM) defined under the risk-neutral measure. In contrast to the RND, transition probabilities are not directly quantifiable from option prices but rather need to be estimated from a given RND surface. In a similar vein, we will denote the real-world transition matrix as $F = (f_{ij})$, and we define the ratio of risk-neutral to real-world transition probabilities as

$$\psi_{ij} = \frac{p_{ij}}{f_{ij}}. \quad (5)$$

3.1.3 This ratio is referred to as the pricing kernel in economics literature (Ross, 1976), the stochastic discount factor in financial economics literature (Cochrane, 2001), and the Radon–Nikodym derivative in option pricing literature (Shreve, 2004). Regardless of its name, $\psi_{ij} > 0$ represents the factor that transforms risk-neutral transition probabilities into their real-world counterparts.¹⁰ This also mathematically illustrates the point made earlier in Section 2: namely, that a change in risk-neutral probabilities does not automatically imply a change in real-world probabilities.

3.1.4 Equation 5 also makes it clear that one needs to solve for two unknowns simultaneously in order to recover the real-world probabilities. In order to do this, we start by assuming that the pricing kernel is transition-independent (i.e., independent of the asset path). This assumption allows us to then define the pricing kernel as

$$\psi_{ij} = \delta \frac{h(S_j)}{h(S_i)}, \quad (6)$$

where $h(S_i)$ is a positive function of the states and δ is a positive discount factor. Combining equations 5 and 6, we have

$$p_{ij} = \delta \frac{h(S_j)}{h(S_i)} f_{ij}. \quad (7)$$

¹⁰ The pricing kernel must be positive to ensure no arbitrage (Ross, 1976).

3.1.5 Recovery of the real-world probabilities thus relies on estimating the values for P , δ and h from the option-implied RND only, which at first glance appears impossible. However, by imposing certain constraints on the matrix P , Ross (2015) shows that this can in fact be achieved. In particular, if one assumes that P is non-negative, irreducible and time-homogeneous, then according to the Perron–Frobenius theorem there exists a unique positive eigenvalue value λ and corresponding unique positive eigenvector z such that

$$Pz = \lambda z. \tag{8}$$

3.1.6 Letting $H = \text{diag}(h(S_1), h(S_2), \dots, h(S_n))$, we can rewrite equation 7 in matrix notation

$$P = \delta H^{-1} F H \Leftrightarrow F = \frac{1}{\delta} H P H^{-1}. \tag{9}$$

Using this expression for F and the fact that each row of this real-world TPM must sum to one, we can write

$$\mathbf{1} = F \mathbf{1} = \left(\frac{1}{\delta} H P H^{-1} \right) \mathbf{1}, \tag{10}$$

where $\mathbf{1}$ is an n -vector of ones. Finally, we rearrange equation 10 to obtain

$$P(H^{-1}\mathbf{1}) = \delta(H^{-1}\mathbf{1}). \tag{11}$$

3.1.7 Written in this form, it becomes clear that equations 11 and 8 are equivalent if and only if $z = H^{-1}\mathbf{1}$ and $\lambda = \delta$. What this means practically is that one can obtain all three unknown variables in equation 7 directly from the option-implied P matrix using an eigenvalue decomposition and thus successfully recover the real-world density. Ross (2015) formalises this result in his recovery theorem:

If the market is arbitrage-free, if the pricing matrix is irreducible and if it is generated by a transition-independent kernel, then there exists a unique (positive) solution to the problem of finding the natural probability transition matrix, F , the discount factor, δ , and the pricing kernel, ψ .

3.2 Implementing the Recovery Theorem

3.2.1 To date, there have only been a handful of empirical studies on the recovery theorem (Spears (2013), Audrino et al. (2015), Backwell (2015), Kiriu & Hibiki (2015), Tran & Xia (2015), and Jackwerth & Menner (2017)). A common observation across these studies is that it is very difficult to implement this theorem. The reason for this is because successful recovery requires one to solve two ill-posed problems. The first of these—estimating the RND surface—has been discussed at length in Section 2. The second

problem is the estimation of the risk-neutral TPM from the obtained RND surface. In contrast to the vast range of RND literature, to our knowledge only the six papers noted above have considered this secondary problem in any level of detail. Given that estimation of the TPM plays such a crucial role in the practical implementation of the recovery theorem, we spend some time below discussing the various aspects of the estimation procedure.

3.2.2 The initial estimation method put forward by Ross (2015) makes use of the assumption of a time-homogeneous TPM to set up a system of linear equations

$$Q'_{1,n,\tau} P = Q'_{1,n,\tau+1}, \quad (12)$$

where $Q'_{1,n,\tau}$ denotes the discretised RND of term τ across the specified n states in P . Equation 12 means that the RND of term $\tau+1$ is equivalent to the product of the RND at term τ and the constant TPM. Tran & Xia (2015) show that the state discretisation specified for the P and Q matrices can materially alter the recovered probability values in certain settings, meaning that setting the state space should be viewed as an active decision in the recovery process. However, they also show that recovery can be consistent across differing state specifications provided that the varying P matrices are consistent in terms of the sum of smaller discretised states adding up to the equivalent larger discretised states.¹¹

3.2.3 Letting $A = Q_{1,n,1:T-1}$ and $B = Q_{1,n,2:T}$, we can write equation 12 in a standard ordinary least squares (OLS) form,

$$P = \underset{p_{ij} \geq 0}{\operatorname{argmin}} \| AP - B \|_2^2. \quad (13)$$

The A and B matrices can be quite large in practice, making direct optimisation of this objective function an onerous exercise. Thankfully, one can recast the problem as a series of independent vector OLS problems which can be solved much more easily and quickly by standard optimisation packages.

3.2.4 In theory then, it would seem that the second ill-posed estimation problem has a fairly simple solution. However, when Spears (2013) attempted to replicate the results originally presented by Ross (2015) using the estimation method given above, the replication was considerably different to the original. This suggests that Ross included additional constraints on the structure of the transition matrix. To this end, Spears (2013) tested nine alternative constrained estimation methods and, using a range of fitting criteria, found that one needs to impose considerable structure on the transition matrix in order to obtain a solution which is both economically suitable and statistically robust.

3.2.5 Audrino et al. (2015) consider the alternative route of using Tikhonov regularisation on the constrained OLS problem (Tikhonov & Arsenin, 1977) rather than impose constraints directly on the transition probabilities. In essence, the idea of regularisation

¹¹ Although not shown below, we test several alternative state space grids in our algorithm below and find little difference in the recovered results.

is to introduce an additional term into the objective function which penalises the optimisation from estimating a P matrix that is too far away from a predefined target matrix. Classic Tikhonov regularisation uses the null matrix as the target but one can generalise this to any target matrix depending on the type of structure that one wants to impose.

3.2.6 Kiriu & Hibiki (2015) select a target transition matrix $\bar{P} = f(Q)$ constructed from the input Q matrix that ensures that the highest transition probabilities are generally found along the diagonal (see Appendix B for construction details). This means that one is assuming that the underlying is more likely to remain in its current state than move to a new state. Kiriu & Hibiki thus attempt to solve the following regularised OLS problem:

$$P = \underset{P_{ij} \geq 0}{\operatorname{argmin}} \|AP - B\|_2^2 + \zeta \|P - \bar{P}\|_2^2, \tag{14}$$

where $\zeta > 0$ is the regularisation parameter that governs the weight given to the additional regularisation norm. Setting $\zeta = 0$ returns the original OLS problem. Although regularisation techniques are often used to very good effect to stabilise the solution set in ill-posed problems, they do introduce the additional issue of selecting the optimal regularisation value, ζ^* . In order to find this optimal value, one needs to introduce a new function that measures the trade-off between solution smoothness and target distance. Common examples include functions based on Euclidean distance (Backwell, 2015), relative entropy (Audrino et al., 2015) or problem-specific selection functions (Kiriu & Hibiki, 2015). After testing each of the respective methods proposed in the above papers, we choose to adopt the selection function proposed by Kiriu & Hibiki (2015) for its appreciably better robustness. See Appendix B for function details.

3.2.7 Having outlined the necessary theoretical and practical issues, we implement the recovery theorem as follows:

- Estimate standardised RNDs as per the procedure given in ¶2.3 and construct a discrete 21-state Q matrix spanning a 50–150% range of the prevailing spot level in 5% intervals.¹²
- Set the TPM period length as three months, in line with the underlying market expiry structure. The input matrices in the OLS problem are thus defined as $A = Q_{1:21,1:T-3}$ and $B = Q_{1:21,4:T}$.
- Construct Kiriu & Hibiki’s (2015) regularisation target matrix \bar{P} and solve the regularised OLS problem in equation 14 using the constrained linear least-squares solver in MATLAB® for a wide range of regularisation values, $\zeta = \{0, 10^{-5}, 10^{-4.8}, \dots, 10^2\}$.
- Find the regularisation value that minimises the selection function, $\zeta^* = \min[h(\zeta)]$.
- Use the optimal ζ^* to solve another regularised OLS problem on a finer 51-state Q matrix in order to estimate a final risk-neutral transition probability matrix P^* .

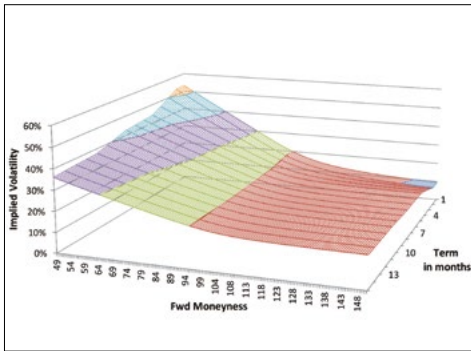
¹² We find consistent recovery results for Q matrices ranging between 21 and 51 states. The lower bound is chosen to reduce computation time of the regularised OLS problems over the full data sample.

— Use the estimated P^* matrix and the recovery theorem to obtain the three-month real-world transition probability matrix F by applying the Perron–Frobenius theorem, and thereafter extract the discrete three-month real-world return distribution as the middle row of the F matrix.

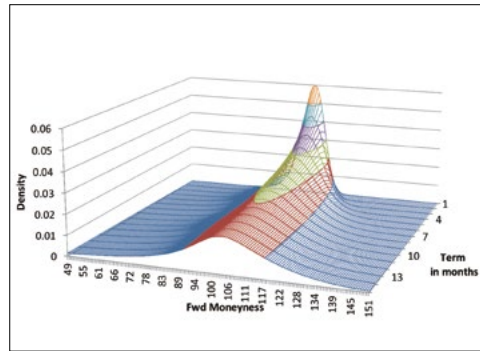
3.2.8 Figure 1 displays the complete recovery procedure based on option trade data as at 23 April 2007. Notice the difference in three-month mean estimates under the implied risk-neutral and real-world distributions.

3.3 Limitations of and Extensions to the Recovery Theorem

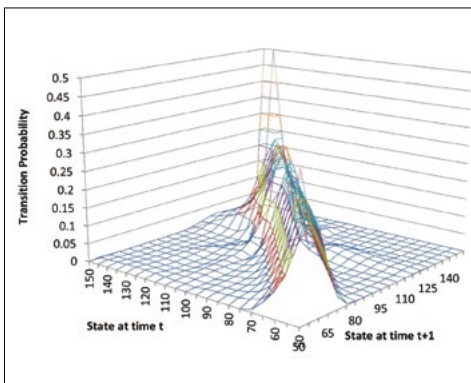
3.3.1 The Ross (2015) recovery theorem as defined and implemented above relies on several fairly stringent market assumptions that are generally violated empirically. Firstly, the market is assumed to be complete and arbitrage-free, and the underlying is constrained to take on only a finite number of states. The first two assumptions are fairly



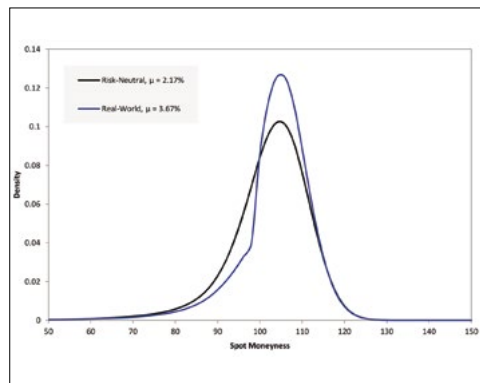
(a) Fitted SVI volatility surface



(b) Risk-neutral distribution surface



(c) Transition probability matrix



(d) Three-month implied distributions

FIGURE 1. Representative diagram of the real-world probability recovery process for options data from 23 April 2007

standard in the majority of derivative pricing models. In terms of recovery, the assumption of complete markets implies that the upward shift of the implied volatility surface due to non-zero transaction costs is not accounted for. The assumption of finite underlying prices is fairly benign and has been relaxed by Carr & Yu (2012).

3.3.2 Secondly, the market is assumed to follow a discrete-time Markov process with a time-homogeneous, irreducible, non-negative transition probability matrix. These are perhaps the most stringent conditions required by the theorem. In terms of the asset process assumptions, the original recovery theorem has since been extended to the more general setting of unbounded, continuous-time Markov processes by Carr & Yu (2012), Walden (2014), Qin & Linetsky (2015) and Park (2016). The assumption that, arguably, has been challenged the most by academics and practitioners is that of time-homogeneous transition probabilities. This means that the probability of moving from a given level today to another level in the future remains constant regardless of changes in future market characteristics. It is well known that this condition is not met in practice and this violation has led some to question the practical value of the theorem (see, for example, Borovička et al. (2013)). However, several researchers have attempted to tackle these criticisms directly by proposing altered forms of the recovery theorem. In this vein, Jensen et al. (2016) and Jackwerth & Menner (2017) have very recently introduced generalised forms of the recovery theorem which do not require time-homogeneity or Markovian processes. The results of Jackwerth & Menner (2017) show that the generalised recovered distribution performs similarly to an economically constrained recovery distribution.

3.3.3 The recovery theorem also assumes that the pricing kernel scales by a constant across periods and therefore it is path-independent. This essentially means that an investor's risk aversion for a given market level would be equivalent even if the market had previously spiked up to that level or alternatively crashed down to that level. This assumption is violated in practice and, furthermore, does not admit the usual suite of CARA or CRRA utility functions.¹³ However, Jensen et al. (2016) have shown that it is possible to adapt the recovery theorem to allow for specific utility functions such as CRRA. As mentioned in Section 1 and in line with Audrino et al. (2015), we take the pragmatic view in this research that unrealistic assumptions do not necessarily invalidate the usefulness of a theorem, provided that the output is of practical value relative to similar output from different models. As a result, we leave further discussion and implementation in illiquid markets of any generalised recovery theorem for future research.

4. TOP40 OPTION-IMPLIED DISTRIBUTIONS

4.1 We use the estimation algorithm outlined in ¶2.3 to create weekly arbitrage-free implied volatility and implied RND surfaces for the Top40 index over the period 5 September 2005 to 16 May 2016, giving a total of 559 surface observations. We then estimate three-

¹³ The acronyms respectively stand for Constant Absolute Risk Aversion and Constant Relative Risk Aversion.

month implied real-world distributions using the recovery algorithm given in Section 3.2. Similarly to Audrino et al. (2015), we consider the evolution and correlation of the first four implied moments—mean, volatility, skewness and kurtosis—relative to the underlying asset over the test period and use these moments as input signals in an index/cash timing strategy.

4.2 Before analysing these moments, we consider the general evolution of the Top40 implied distribution. Figure 2 displays a dynamic boxplot of the three-month implied distribution over the ten-year sample period. This information can be used descriptively or prescriptively. In the descriptive sense, the negative tail of the distribution (red shading) is consistently longer than the positive tail (green shading) throughout the period, indicative of the larger negative jump or crash risk common to equity indices. The ratio of these two areas thus describes the level of implied asymmetry in the index at any point in time. The RND widened significantly during the global financial crisis and remained wider than usual until late 2009. Since then, the RND has narrowed considerably, although the negative tail has once again started to increase over the last couple of observed years. In the prescriptive sense, one could for example focus on the 5th distribution percentile. This is essentially the quarterly implied value-at-risk of the Top40 index and can thus be used in a number of forward-looking risk management applications.

4.3 Risk-Neutral and Real-World Implied Moments

4.3.1 We now consider the implied moments of the risk-neutral and recovered three-month distributions. Figure 3 overleaf compares the first four risk-neutral moments to their real-world counterparts, against the performance backdrop of the underlying Top40 total return index. Table 1 below provides the corresponding summary statistics. Notice that the real-world mean is almost always considerably higher than its risk-neutral counterpart—essentially the quarterly cost of carry—and also displays significantly more time variation. Table 1 also shows that the annualised real-world mean (and volatility) is very close to the

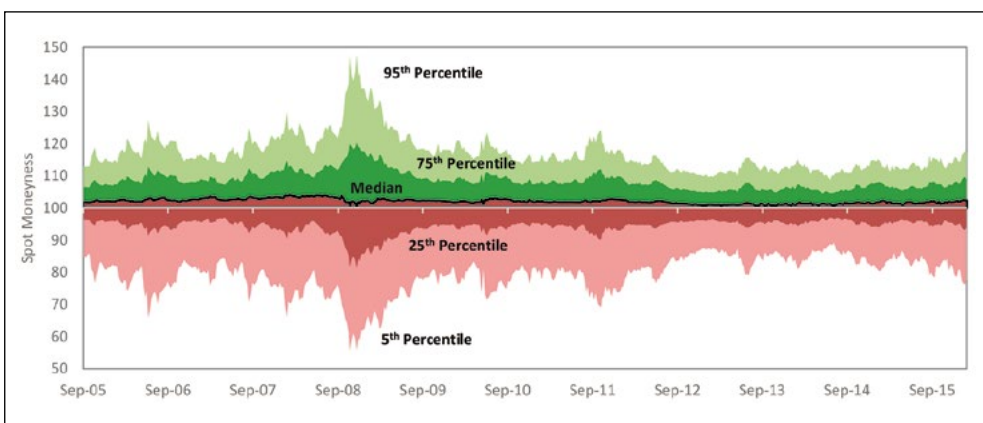


FIGURE 2. Weekly box plot of three-month risk-neutral Top40 distributions
September 2005–January 2016

historical annualised Top40 return mean (and volatility) over the sample period. We stress again that recovery of a real-world mean estimate purely from the options market is a remarkable feat and gives one considerable insight into the actual market views used by option market participants for pricing purposes. Coupled with a greater macro view of the derivatives market, this also gives one an inkling of how structural issues such as liquidity and the supply/demand ratio may affect the implied market views used in derivatives pricing.

TABLE 1. Implied moment summary statistics, September 2005–May 2016

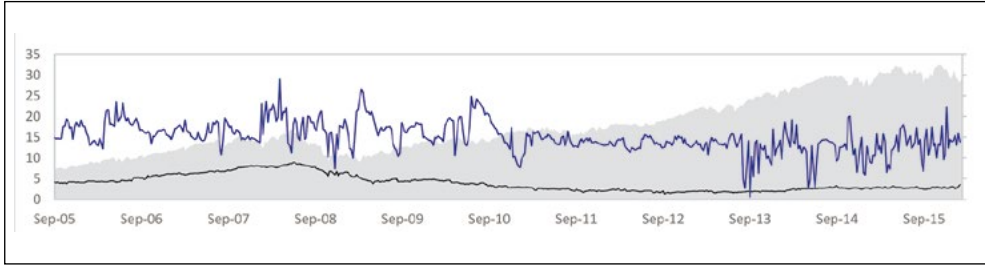
	Top40 returns ¹	Risk-neutral				Real-world			
		Mean ²	Volatility ²	Skewness	Kurtosis	Mean ²	Volatility ²	Skewness	Kurtosis
Mean	16.10%	4.0%	23.5%	-0.91	4.95	15.1%	21.3%	-1.05	5.39
Volatility	21.50%	1.9%	6.8%	0.29	1.24	3.6%	6.6%	0.28	1.34
Minimum	-11.70%	1.3%	12.4%	-1.67	2.68	0.6%	10.3%	-1.71	2.21
5th	-4.70%	1.8%	15.2%	-1.40	3.20	9.1%	13.1%	-1.48	3.31
25th	-1.20%	2.4%	18.8%	-1.12	4.16	13.1%	17.0%	-1.25	4.53
50th	0.50%	3.3%	22.1%	-0.92	4.75	14.7%	19.9%	-1.07	5.38
75th	2.10%	4.9%	26.6%	-0.74	5.62	17.4%	24.2%	-0.87	6.26
95th	4.60%	7.9%	34.7%	-0.45	7.21	21.2%	32.1%	-0.62	7.67
Maximum	15.30%	9.0%	55.8%	0.09	10.97	28.9%	48.3%	-0.08	11.00

¹Top40 Returns mean and volatility is annualised; percentiles are weekly numbers.
²Mean and volatility numbers are all annualised.

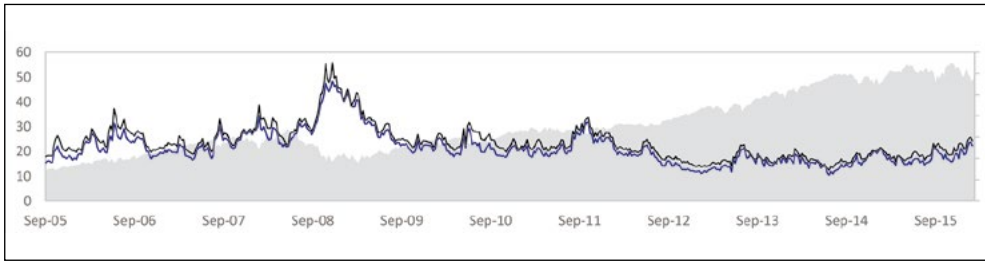
4.3.2 From the second panel in Figure 3, we can clearly observe the well-known inverse relationship between asset performance and implied volatility. More importantly though, the two implied volatility profiles are very similar, in line with what one would expect given the theory underlying the BSM pricing framework after accounting for a potential volatility premium added by market makers. This suggests that our recovery algorithm is giving us results that at least match the minimum arbitrage pricing criteria.

4.3.3 The recovered skewness is almost always lower than the risk-neutral skewness, although both are considerably negative as one would expect for three-month index returns. Furthermore, and in line with the literature, both implied distributions are essentially symmetric during the global financial crisis and only show significant left-skew post the market recovery. A similar but inverted relationship is seen between implied kurtosis and index performance during this period. However, there are also several periods in the full ten-year sample when kurtosis declines but the index displays positive performance, making it difficult to generalise this finding without further analysis.

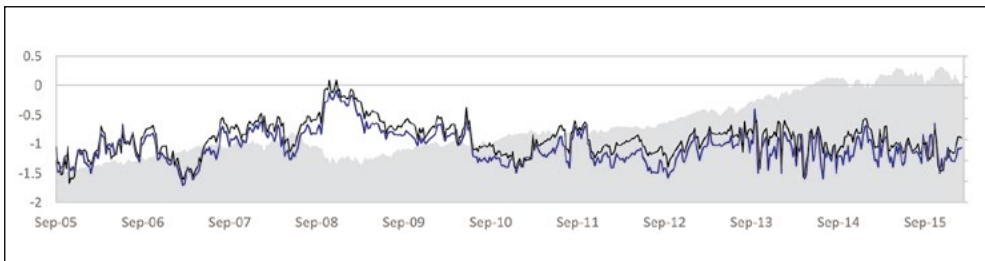
4.3.4 Table 2 displays the correlation matrix for changes in the weekly risk-neutral and real-world implied moments. The table has been split into quadrants, which, reading anti-clockwise from top left, denote correlations between risk-neutral moments only,



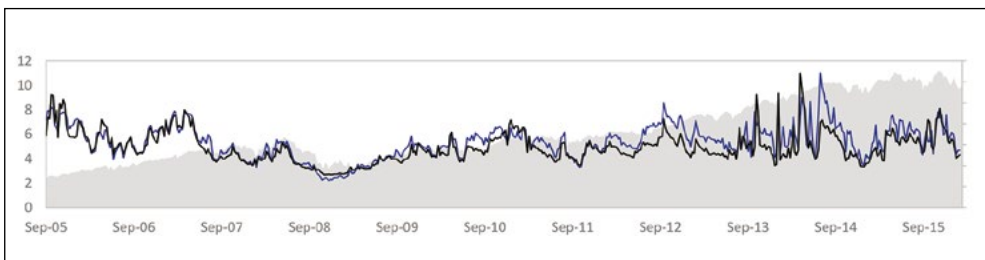
(a) Annualised mean, %



(b) Annualised volatility, %



(c) Three-month skewness



(d) Three-month kurtosis

FIGURE 3. Top40 real-world vs risk-neutral moments, with index total return performance as backdrop

correlations between risk-neutral and real-world moments, and correlations between real-world moments only. Comparing the upper left and lower right quadrants, one observes that the real-world higher moments have considerably stronger lower moment relationships than their respective risk-neutral counterparts. This is particularly noticeable for kurtosis.

TABLE 2. Correlation matrix of weekly implied moment changes, September 2005–May 2016

		Risk-neutral				Real-world			
Top40		Mean	Volatility	Skewness	Kurtosis	Mean	Volatility	Skewness	Kurtosis
Risk-neutral	Mean	1.000							
	Volatility	-0.321	1.000						
	Skewness	-0.153	0.243	1.000					
	Kurtosis	0.090	-0.131	-0.812	1.000				
Real-world	Mean	0.057	-0.019	-0.115	-0.075	1.000			
	Volatility	-0.230	0.880	0.302	-0.144	-0.331	1.000		
	Skewness	-0.133	0.294	0.731	-0.397	-0.354	0.392	1.000	
	Kurtosis	0.085	-0.304	-0.677	0.585	0.460	-0.513	-0.791	1.000

4.3.5 The lower left quadrant reveals the strong positive relationship between risk-neutral and real-world volatility as well as significantly positive correlations between the two skewness and kurtosis measures respectively. However, there is still evidence to suggest that the informational content available from each pair of moments is different. This is particularly evident when observing the large differences in the correlations between the real-world mean and the other risk-neutral moments versus the correlations to the other real-world moments.

4.4 Tactical Asset Allocation with Option-Implied Information

4.4.1 We heuristically test the forward-looking information content of the moments by following a simple tactical asset allocation (TAA) strategy over the sample period, as advocated by Audrino et al. (2015). For expected return, skewness and kurtosis, if the current week’s values are greater than the prior week’s, then we hold the Top40 index, otherwise we move into cash.¹⁴ We take the opposite strategy for volatility given the well-known inverse relationship with underlying returns. Although simple, this strategy is in line with an investor wanting higher returns, higher skewness and higher kurtosis.¹⁵

¹⁴ Transaction costs are not included as we are only interested in assessing the informational content for now. Furthermore, the number of trades is fairly consistent across all timed strategies, meaning that costs would have a similar impact throughout.

¹⁵ Note that an investor will generally only want higher kurtosis, and thus fatter tails, when skewness becomes increasingly positive.

4.4.2 Figure 4 displays the cumulative log-returns of the strategies versus the Top40 total return in black. The blue shaded lines are the risk-neutral strategies while the red shaded lines are the real-world strategies. Table 3 gives the summary statistics from the trading strategies.

4.4.3 The most striking observation from the depicted and tabulated results is that the recovered moment strategies consistently and considerably outperform the Top40 total return index and the risk-neutral moment strategies, with the exception of the risk-neutral skewness strategy. The average return for the real-world strategies ranges between 16.9% and 18.1%, which is 0.8% to 1.9% higher than the Top40 return over the same period. The volatility of the real-world strategies is also considerably lower than that of the Top40, meaning that the risk-adjusted performance of the physical index—as measured by the Sharpe ratio—is consistently lower than the real-world timed strategies. Interestingly, the risk-adjusted comparison between risk-neutral and real-world strategies is not as clear cut. The mean and volatility strategies are clearly dominated by the real-world moments, whereas the comparison is much closer for the higher moment strategies. This suggests two points: firstly, that higher moments are important in a TAA context, and secondly, that the information content within the implied risk-neutral higher moments may be as valuable as that gained from the recovered real-world counterparts. We leave a more detailed discussion of this conjecture for later research.

4.4.4 In Table 3 the large deviation in means between the risk-neutral and real-world volatility strategies is of interest, particularly in light of the high correlation between the two underlying volatility series. To understand this, one must remember that the signals in these tactical strategies are based only on the sign and not the size of the changes in the respective volatility series. Looking at the differences in the binary indicators created from these two volatility series, we find that the strategies hold different market positions for 15% of the full test period, which is significant for a tactical timing strategy. It is this difference in allocations that drives the difference in overall strategy performance.

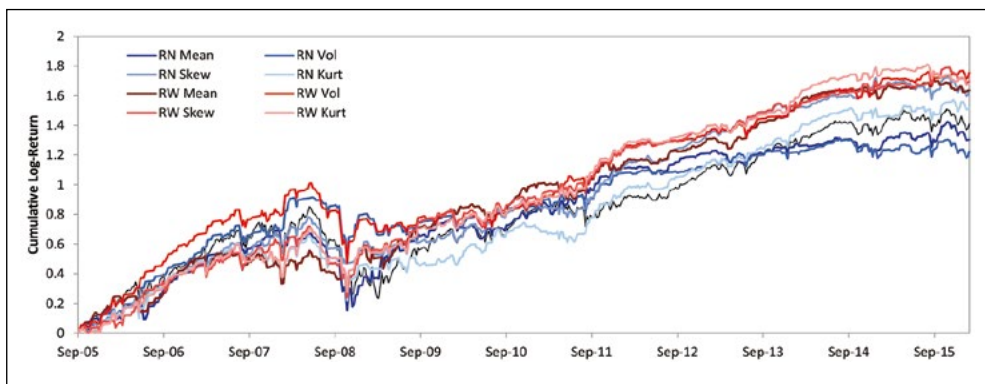


FIGURE 4. Cumulative log performance of implied trading strategies versus Top40 index, September 2005–May 2016

TABLE 3. Top40 implied moment trading strategy results, September 2005–May 2016

	Top 40	Risk-neutral moment TAA				Real-world moment TAA			
		Mean	Volatility	Skewness	Kurtosis	Mean	Volatility	Skewness	Kurtosis
Mean	16.12%	14.25%	13.82%	17.03%	16.67%	16.94%	18.05%	17.51%	17.08%
Volatility	21.71%	17.14%	16.36%	15.25%	16.66%	15.83%	16.02%	15.84%	16.19%
Sharpe Ratio	0.50	0.52	0.52	0.77	0.69	0.74	0.80	0.77	0.73
Skewness	-0.24	-0.13	0.07	-0.14	0.13	0.28	0.03	0.35	0.23
Kurtosis	5.99	11.37	12.01	8.19	10.41	12.45	12.97	12.17	11.51
No. Trades	n/a	289	243	283	268	296	269	293	291
Max DD ¹	-46.17%	-41.54%	-33.56%	-26.78%	-35.18%	-26.97%	-41.57%	-38.04%	-34.77%

¹ Max DD = Maximum drawdown

4.4.5 Kurtosis of the timed strategy returns is significantly higher than for the index. However, skewness is generally positive as well, meaning that the timed strategies actually display significant positive tail risk. This stems from the fact that one generally moves to cash during some of the worst market downturns, thus decreasing the number and size of the negative tail events. This can also be seen in the reduced maximum drawdown numbers relative to the index portfolio.

4.4.6 In summary, the heuristic information testing of the implied moments would suggest that there is merit in recovering the real-world moments, at least in the case of tactical asset allocation.

5. CONCLUSION

5.1 Given the forward-looking nature of the derivatives market, it is reasonable to surmise that there may be information embedded in option market prices. Numerous authors have shown that such option-implied information significantly outperforms the comparative information estimated from price history across a range of portfolio, risk management and trading applications. Although the estimation of risk-neutral option-implied information is well-established in the literature, estimation of the same in an illiquid market is not. Furthermore, there has been little empirical research done to date—in liquid and illiquid markets alike—on extracting real-world implied information using the recovery theorem introduced by Ross (2015). In this work, we address both these issues by considering in detail the estimation and application of risk-neutral and real-world option-implied distributions in an illiquid market setting.

5.2 We show that the deterministic SVI volatility model is a viable candidate for modeling implied volatility surfaces and use this model to estimate the underlying risk-neutral

distributional surfaces on the Top40 index. The issue of calibration with sparse and noisy data is considered and a simple but robust fitting algorithm is proposed.

5.3 We then describe a robust methodology based on regularised least squares for extracting these implied real-world probabilities and implement this method on a history of weekly SVI implied volatility surfaces for the Top40 index. We discuss how one can use this information descriptively and prescriptively and, furthermore, analyse the recovered moments from the implied distributions. The recovered real-world moments are shown to be in line with economic rationale and also show promising results when used as signals within a simple tactical asset allocation framework.

5.4 Potential avenues for further research include the implementation of generalised forms of the recovery theorem in illiquid markets. In particular, it would be interesting to compare how various versions of semi-parametric recovered distributions, either with constraints on possible recovered distributions or possible recovered pricing kernels, compare to the standard recovered information. Furthermore, it would be interesting to consider further applications of recovered real-world information in the portfolio and risk management space in line with similar work in this field done using risk-neutral implied information.

ACKNOWLEDGEMENTS

We are grateful for discussions with Anthony Seymour and Florence Chikurunhe and for the constructive feedback from the 2016 ASSA Convention and SAAJ reviewers respectively. We are also grateful to the SAAJ editorial team for their help during the submission process.

AUTHOR DECLARATION

We declare no conflicts of interest in publishing this article and welcome any and all comments. The views expressed here are only our own and any errors are our responsibility.

REFERENCES

- Aït-Sahalia, Y & Duarte, J (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics* **116**(1), 9–47
- Aït-Sahalia, Y & Lo, AW (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance* **53**(2), 499–547
- Audrino, F, Huitema, R, & Ludwig, M (2015). An empirical analysis of the Ross recovery theorem. Available at SSRN 2433170
- Backwell, A (2015). State prices and implementation of the recovery theorem. *Journal of Risk and Financial Management* **8**(1), 2–16
- Bates, DS (2000). Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics* **94**(1), 181–238
- Baule, R, Korn, O & Saßning, S (2015). Which beta is best? On the information content of option-implied betas. *European Financial Management* **22**(3), 450–83
- Black, F (1976). The pricing of commodity contracts. *Journal of Financial Economics* **3**(1), 167–79
- Black, F & Scholes, M (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy* **81**(3), 637–54
- Bliss, RR & Panigirtzoglou, N (2002). Testing the stability of implied probability density functions. *Journal of Banking & Finance* **26**(2), 381–422
- Bliss, RR & Panigirtzoglou, N (2004). Option-implied risk aversion estimates. *The Journal of Finance* **59**(1), 407–46
- Borovička, J, Hansen, LP & Scheinkman, JA (2016). Misspecified recovery. *The Journal of Finance* **71**(6), 2493–544
- Breeden, DT & Litzenberger, RH (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business* **51**(4) (Oct 1978), 621–51
- Buchen, PW & Kelly, M (1996). The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis* **31**(01), 143–59
- Buss, A & Vilkov, G (2012). Measuring Equity risk with option-implied correlations. *Review of Financial Studies* **25**(10), 3113–40
- Carr, P, Geman, H, Madan, DB & Yor, M (2003). Stochastic volatility for Lévy processes. *Mathematical Finance* **13**(3), 345–82
- Carr, P & Madan, DB (2005). A note on sufficient conditions for no arbitrage. *Finance Research Letters* **2**(3), 125–30
- Carr, P & Yu, J (2012). Risk, return, and Ross recovery. *Journal of Derivatives* **20**(1), 38
- Cochrane, JH (2001). *Asset Pricing*. Princeton University Press.
- Cooper, N (1999). Testing techniques for estimating implied RNDs from the prices of European-style options. In Proceedings of the BIS workshop on implied PDFs, Bank of International Settlements, Basel
- Coutant, S, Jondeau, E & Rockinger, M (1998). Reading interest rate and bond futures options' smiles: How PIBOR and notional operators appreciated the 1997 French snap election. Technical report, Banque de France
- Cox, JC & Ross, SA (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics* **3**(1–2), 145–66

- DeMarco, P & Martini, C (2009). *Quasi-Explicit Calibration of Gatheral SVI model*. Zeliade Systems.
- DeMiguel, V, Garlappi, L & Uppal, R (2009). Optimal Versus naïve diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies* **22**(5), 1915–53
- DeMiguel, V, Plyakha, Y, Uppal, R & Vilkov, G (2013). Improving portfolio selection using option-implied volatility and skewness. *Journal of Quantitative & Financial Analysis* **48**(6), 1813–45
- Dubynskiy, S & Goldstein, RS (2013). Recovering drifts and preference parameters from financial derivatives. Available at SSRN 2244394.
- Dumas, B, Fleming, J & Whaley, RE (1998). Implied volatility functions: Empirical tests. *The Journal of Finance* **53**(6), 2059–106
- Figlewski, S (2008). Estimating the Implied risk neutral density. volatility and time series econometrics: Essay in honor of Robert F. Engle, Tim Bollerslev, Jeffrey R. Russell, Mark Watson, Oxford, UK: Oxford University Press
- Gatheral, J (2004). A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives. Presentation at Global Derivatives & Risk Management, Madrid
- Gatheral, J (2011). *The Volatility Surface: A Practitioner's Guide*, Volume 357. John Wiley & Sons
- Gatheral, J & Jacquier, A (2014). Arbitrage-free SVI volatility surfaces. *Quantitative Finance* **14**(1), 59–71
- Hagan, PS, Kumar, D, Lesniewski, AS & Woodward, DE (2002). Managing smile risk. *Wilmott Magazine* (July 2002) 84–108
- Heston, SL (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* **6**(2), 327–43
- Jackwerth, JC & Menner, M (2017). Does the Ross recovery theorem work empirically? Available at SSRN 2960733
- Jensen, CS, Lando, D & Pedersen, LH (2016). Generalized recovery. Available at SSRN 2674541
- Kempf, A, Korn, O & Saßning, S (2015). Portfolio optimization using forward-looking information. *Review of Finance* **19**(1), 467–90
- Kiritu, T & Hibiki, N (2015). Estimating forward looking distribution with the Ross recovery theorem. Keio University Working Paper
- Kostakis, A, Panigirtzoglou, N & Skiadopoulos, G (2011). Market timing with option-implied distributions: A forward-looking approach. *Management Science* **57**(7), 1231–49
- Kotzé, A & Joseph, A (2009). Constructing a South African index volatility surface from exchange traded data. Available at SSRN 2198357
- Qin, L & Linetsky, V, 2015. Positive eigenfunctions of Markovian pricing operators: Hansen-Scheinkman factorization, Ross recovery and long-term pricing. Available at SSRN 2439002
- Malz, AM (2014). A simple and reliable way to compute option-based risk-neutral distributions. Technical Report 677, FRB of New York
- Martin, I & Ross, S (2013). The long bond. Technical report, London School of Economics and MIT
- McManus, DJ (1999). The information content of interest rate futures options. Working Paper 99–15, Bank of Canada

- Melick, WR & Thomas, CP (1997). Recovering an asset's implied PDF from option prices: An Application to crude oil during the Gulf crisis. *Journal of Financial and Quantitative Analysis* **32**(01), 91–115
- Merton, RC (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science* **4**(1), 141–83
- Merton, RC (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* **3**(1–2), 125–44
- Park, H (2016). Ross recovery with recurrent and transient processes. *Quantitative Finance* **16**(5), 667–76
- Posner, SE & Milevsky, MA (1998). Valuing exotic options by approximating the SPD with higher moments. *The Journal of Financial Engineering* **7**(2), 109–25
- Roper, M (2010). Arbitrage free implied volatility surfaces. Technical report, University of Sydney Working Paper
- Ross, S (2015). The recovery theorem. *The Journal of Finance* **70**(2), 615–48
- Ross, SA (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory* **13**(3), 341–60
- Shimko, D (1993). The bounds of probability. *RISK* **6**, 33–7
- Shreve, SE (2004). *Stochastic Calculus for Finance II: Continuous-time Models*, Volume 11. Springer Science & Business Media
- Spears, T (2013). On Estimating the risk-neutral and real-world probability measures. Master's thesis, Oxford University
- Tikhonov, A & Arsenin, VY (1977). *Methods for Solving Ill-Posed Problems*. John Wiley and Sons, Inc.
- Tompkins, RG (2001). Implied volatility surfaces: Uncovering regularities for options on financial futures. *The European Journal of Finance* **7**(3), 198–230
- Tran, N-K & Xia, S (2015). Specified recovery. Technical report, Washington University in St. Louis
- Walden, J (2014). Recovery with unbounded diffusion processes. Available at SSRN 2508414
- West, G (2005). Calibration of the SABR Model in illiquid markets. *Applied Mathematical Finance* **12**(4), 371–85

APPENDIX A

SVI Calibration Algorithm

A.1 We summarise the SVI calibration procedure given in DeMarco & Martini (2009). Define $w(x) = \tau\sigma^2(x)$ as a single total variance skew and introduce the change of variables

$$y = \frac{x - m}{s}. \quad (15)$$

A.2 Using this new variable y , we can rewrite the SVI equation in total variance space as

$$w(y) = \alpha + \delta y + \beta\sqrt{y^2 - 1}, \quad (16)$$

where

$$\begin{aligned} \alpha &= a\tau \\ \beta &= bs\tau \\ \delta &= \rho bs\tau. \end{aligned} \quad (17)$$

A.3 Taking m and s as fixed values, the objective function then becomes

$$\operatorname{argmin}_{(\alpha, \beta, \delta) \in D} \sum_{i=1}^n w_i \left(\alpha + \delta y + \beta\sqrt{y^2 - 1} - V_i \right)^2, \quad (18)$$

where n is the total number of market observations for the given expiry and $V_i = v_i\tau$ is the i^{th} observed total variance. The domain D is defined as

$$D = \left\{ \begin{array}{l} 0 \leq \beta \leq 4s \\ |\delta| \leq \beta \text{ and } |\delta| \leq 4s - \beta \\ 0 \leq \alpha \leq \max\{v_i\}. \end{array} \right. \quad (19)$$

A.4 For a given solution set $\{\alpha^*, \beta^*, \delta^*\}$, one can find the corresponding solution set $\{a^*, b^*, \rho^*\}$ and thus the remaining two parameters can be calibrated as follows:

$$\operatorname{argmin}_{m, s \geq 0} \sum_{i=1}^n w_i \left(\sigma^2 \left(x, \{m, s, a^*, b^*, \rho^*\} \right) - v_i \right)^2. \quad (20)$$

The original five-dimensional calibration is thus broken into separate three-dimensional and two-dimensional minimisation problems.

APPENDIX B

Regularisation Parameters

B.1 Kiriu & Hibiki (2015) calculate the regularisation target matrix \bar{P} directly from the discretised RND matrix Q based on two premises. Firstly, because the expiry in the first column of Q is equal to the expiry of the transition probability matrix P , and because the states are chosen symmetrically around the current market level, it must be that the middle row of the P matrix is equal to the first column of Q . Secondly, Kiriu & Hibiki (2015) suggest that the probability of transitioning from states S_i to S_j should be similar to the probability of transitioning from states S_{i+k} to S_{j+k} for all $k < \min(n-i, n-j)$. The first premise defines the middle row of \bar{P} while the second defines the remainder of the matrix. Assuming that one has an odd number n of states and defining $m_0 = (n+1)/2$ as the middle state, we have

$$\bar{P} = \begin{bmatrix} \sum_{i=1}^{m_0} q_{i,1} & q_{m_0+1,1} & \cdots & q_{n,1} & \cdots & 0 & 0 \\ \sum_{i=1}^{m_0-1} q_{i,1} & q_{m_0,1} & \cdots & q_{n-1,1} & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots \\ q_{1,1} & q_{2,1} & \cdots & q_{m_0,1} & \cdots & q_{n-1,1} & q_{n,1} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & q_{2,1} & \cdots & q_{m_0,1} & \sum_{i=m_0+1}^n q_{i,1} \\ 0 & 0 & \cdots & q_{1,1} & \cdots & q_{m_0-1,1} & \sum_{i=m_0}^n q_{i,1} \end{bmatrix}. \quad (21)$$

B.2 Although there are a number of standard functions used to evaluate regularisation parameters, Kiriu & Hibiki (2015) suggest a problem-specific selection function that attempts to balance relative gain in the objective function from each term in the regularised OLS minimisation.

$$\begin{aligned} P &= \underset{p_{ij} \geq 0}{\operatorname{argmin}} \|AP - B\|_2^2 + \zeta \|P - \bar{P}\|_2^2 \\ &= \underset{p_{ij} \geq 0}{\operatorname{argmin}} y_{fit}(\zeta) + \zeta y_{reg}(\zeta). \end{aligned} \quad (22)$$

B.3 The selection function is then given as

$$h(\zeta) = \frac{y_{fit}(\zeta) - y_{fit}(0)}{y_{fit}(\infty) - y_{fit}(0)} + \frac{y_{reg}(\zeta) - y_{reg}(\infty)}{y_{reg}(0) - y_{reg}(\infty)}, \quad (23)$$

where the respective denominators represent the maximum spread in each term and the numerator gives the spread achieved for a specified ζ value. The $y_i(0)$ values are solutions from the original OLS problem and $y_{reg}(\infty)$ is set to zero due to the fact that $P \rightarrow \bar{P}$ as $\zeta \rightarrow \infty$. This implies that $h(0) = h(\infty) = 1$. Kiriu & Hibiki (2015) show under simulation that the h function is smooth, continuous and has a single minimum value, and most importantly, that the derivative function h' is very stable around this global minimum, thus making it a very appealing selection function.