Estimating Option-Implied Distributions in Illiquid Markets and Implementing the Ross Recovery Theorem*

Emlyn Flint† & Eben Maré‡

This Version: 04 August 2016

Abstract

We describe how forward-looking information on the statistical properties of an asset can be extracted directly from options market data and how this can be used practically in portfolio management. Although the extraction of a forward-looking risk-neutral distribution is well-established in the literature, the issue of estimation 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. Furthermore, we 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.

---

*We are grateful for discussions with Anthony Seymour and Florence Chikurunhe. We welcome any and all comments. The views expressed here are only our own and any errors are our responsibility.

†University of Pretoria and Peregrine Securities, South Africa. emlyn.flint@gmail.com
‡University of Pretoria, South Africa. eben.mare@up.ac.za
1 Introduction

An important requirement for optimal portfolio construction is an understanding of the future possible returns of the constituent assets. Armed with this understanding, investors can ensure that their chosen combination of assets will lead to a portfolio that is consistent with their 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. An alternate forecasting method is to extract forward-looking information on the statistical properties of an asset directly from options market data.

The seminal work of [Black and Scholes (1973)](http://ssrn.com/abstract=2817080) and [Merton (1973)](http://ssrn.com/abstract=2817080) 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, called the implied volatility. 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.

Implied volatility surfaces in practice differ 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. 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. Therefore, in addition to providing one with a means of 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.

The idea of accessing this embedded information is not new. Implied volatility has long been used as a 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 1990’s, 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.
Breeden and Litzenberger (1978) proved that the forward-looking risk-neutral 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 (i.e. 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 betas and correlations. Kempf et al. (2015); Baule et al. (2015); DeMiguel et al. (2013); Buss and Vilkov (2012) and Kostakis et al. (2011) among others, have shown that such implied moments and statistics significantly outperform their historical counterparts in a range of portfolio, risk management and trading applications.

An important point to remember though 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.

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 over that found in its more easily available risk-neutral counterpart. Although a number of researchers have raised fundamental questions as to whether there can actually be either unique or better information with 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.

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 chosen technique 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 regularized least squares. Empirical results using South African option 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 concludes.

2 Estimating Option-Implied Distributions in Illiquid Markets

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

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

(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 and Litzenberger (1978)]:

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

(2)

In theory, one needs a continuum of option prices 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.

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 and Milevsky (1998)), or using mixtures of lognormal distributions (Melick and Thomas (1997)). Nonparametric forms include the use of entropy-based methods (Buchen and Kelly (1996)) and kernel-based methods (Aït-Sahalia and Lo (1998)).

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

---

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.
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 and 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).

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’. Of the small number of comparative studies done 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. 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.

This becomes even more important in illiquid markets where option data is both sparse and noisy. Because of these two issues, many of the techniques proposed above become unsuitable. Apart from the shape-constrained kernel method of Aıt-Sahalia and 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 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 et al. (1999) notes a similar result for the 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 Generalized Extreme Value distribution tails onto the estimated central portion of the RND. In either case, the researcher is ultimately prespecifying the tail structure of the RND, thus actually making the spline technique somewhat parametric.

There is also another issue that needs to be considered in our work. 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 Aıt-Sahalia and 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 respectively are Gatheral’s (2004) stochastic volatility inspired (SVI) model and Heston’s (1993) stochastic volatility model, both of which are used extensively by academics and practitioners worldwide.

Of these two candidates, 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 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 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.

### 2.1 The South African Options Market

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 volatility calibration challenges faced within this market; in his case within a SABR model context. At the time of his study, over the counter (OTC) structures comprised the majority of SA option activity. However, the subprime crisis in 2008 changed the manner in which investors traded. A renewed interest in regulation and a renewed 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 listed contracts also meant a far larger proportion of the underlying trade data became publicly available.

In this market, participants mostly use derivatives as hedging tools, meaning that open interest is concentrated in put options with strikes below current index levels. These hedging structures are usually short-term. Volumes are thus concentrated in the three closest expiries - i.e. up to nine months - and trade in any term longer than 15 months is extremely rare. The size of such hedges can also sometime dwarf all other trades in a given period, leading to

---

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

4The Top40 comprises the largest 40 South African stocks listed by gross market capitalisation.
an extremely skewed open strike distribution. That being said, total option volumes remain small in comparison to other developed markets. On any given day, the number of trades varies significantly and there could be 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 index levels.

Daily listed Top40 option trade data is freely available from the Johannesburg Stock Exchange (JSE) website 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 data set generally includes trade date and time, futures level, strike, traded volatility, price, option type and volume. However, 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.

2.2 The Stochastic Volatility Inspired Model

The SVI model was disseminated by Gatheral (2004) and has since 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 \( \tau \) and the strike as \( K \), we can write the SVI implied variance as

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

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 the log-moneyness space. This particular form is chosen because it ensures that variance is linear as \( |x| \to \infty \) - a fundamental characteristic of volatility surfaces - while still being convex around the 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 and Jacquier (2014)):

- \( a \) defines the overall level of variance and shifts the skew vertically;
- \( b \) defines the angle between the put and call wing variance slopes;
- \( \rho \) rotates the variance curve clockwise around the current forward level;
- \( m \) shifts the variance curve left or right;
- \( s \) defines the amount of at the money (ATM) variance curvature.

The smart choice of parameterisation coupled with the five degrees of freedom generally ensures an extremely good fit in practice, particularly in the equity index space. Further-

\footnote{Note that there is no bid-ask spread information included.}
more, 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 and 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.

One drawback of the SVI model though 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 and Martini (2009) addressed this shortcoming though 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. Gatheral and Jacquier (2014) suggest another calibration method based on a similar alternative parameterisation of the original model.

2.3 Constructing SVI Volatility & RND Surfaces

Although international literature on modelling implied volatility is vast, the majority of these studies are not easily applicable to the South African (SA) derivatives market due to its illiquid nature and fairly unique trading dynamics. To the authors’ knowledge, only West (2005) and Kotzé and 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 and Joseph (2009) do the same but using a quadratic deterministic volatility model. Both studies stress the need for robust and sensible calibration algorithms and put forth several useful suggestions for reaching that goal, which are incorporated below. Be that as it may, creating a robust calibration procedure still requires some “creative decision making” as West puts it. In this context, the SVI volatility and RND surface algorithm given here 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:

1. Collate Top40 option trade data for the past seven days. Back fill 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.

6This is the model currently used by SAFEX for mark-to-market and margining purposes.
2. 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 trades falling in prespecified size buckets ($w_i = \{0.1, 0.8, 1, 0.8\}$ for trades of less than 100, 500, 2000 and 10000 contracts respectively).

3. In cases of extreme data sparsity, include several OTM skew markers from the previous period’s calibration, adjusted for the current ATM volatility level.

4. Calibrate the SVI model separately to each traded expiry using the ‘2+3’ algorithm of DeMarco and Martini (2009).

5. 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 and Jacquier (2014).

6. Use the modified SVI parameters to create volatility skews across a 20−300% range of the prevailing forward prices.

7. 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).

8. 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.

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.

3 Recovering Real-World Implied Distributions

Below, we give a brief outline of the Ross recovery theorem along with its underlying assumptions in a style similar to Spears (2013). 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.

7Interested readers can find further mathematical detail in Ross (2015). Extensions to the original theory are presented in Carr and Yu (2012), Dubynskiy and Goldstein (2013), Martin and Ross (2013), Walden (2014) and Borovička, Jaroslav and Hansen, Lars Peter and Scheinkman, José A (2016).
3 RECOVERING REAL-WORLD IMPLIED DISTRIBUTIONS

3.1 The Ross Recovery Theorem

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

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

(4)

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 hold that $p_{ij} > 0$ for some $t$.

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

$$\psi_{ij} = \frac{p_{ij}}{f_{ij}}.$$  

(5)

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.\(^8\) 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.

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

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

(6)

where $h(S_i)$ is a positive function of the states and $\delta$ is a positive discount factor. Combining Equations (5) and (6), we have that

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

(7)

\(^8\)The pricing kernel must be positive to ensure no arbitrage.
Recovery of the real-world probabilities thus relies on estimating the values for \( p \), \( \delta \) 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-independent, then according to the Perron-Frobenius theorem there exists a unique positive eigenvalue value \( \lambda \) and corresponding unique positive eigenvector \( z \) such that

\[
Pz = \lambda z.
\]  
(8)

Letting \( H = \text{diag}(h(S_1), h(S_2), \ldots, h(S_n)) \), we can rewrite Equation 7 in matrix notation,

\[
P = \delta H^{-1}F \iff F = \frac{1}{\delta} HPH^{-1}.
\]  
(9)

Using this expression for \( F \) along with the fact that each row of the real-world TPM must sum to 1, and setting \( \mathbf{1} = (1, \ldots, 1)' \) we can write

\[
\mathbf{1} = F\mathbf{1} = \left( \frac{1}{\delta} HPH^{-1} \right) \mathbf{1}.
\]  
(10)

Finally, we rearrange Equation 10 to obtain

\[
P \left( H^{-1} \mathbf{1} \right) = \delta \left( H^{-1} \mathbf{1} \right).
\]  
(11)

Written in this form, it becomes clear that Equations 11 and 8 are equivalent 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:

\[\text{If the market is arbitrage-free, if the pricing matrix is irreducible and if it is generated by a transition independent kernel, then there exists a unique (positive) solution to the problem of finding the natural probability transition matrix, } F, \text{ the discount factor, } \delta, \text{ and the pricing kernel, } \psi.\]

### 3.2 Implementing the Recovery Theorem

To date, there have only been a handful of empirical studies on the recovery theorem (Spears (2013); Audrino et al. (2015); Backwell (2015); Kiriu and Hibiki (2015) and Tran and Xia (2015)). A common thread running through 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 RND literature, to our knowledge only the five 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.

The initial estimation method put forth by Ross makes use of the assumption of a time-homogeneous TPM to set up a system of linear equations

$$Q_{1,n,t}^t P = Q_{1,n,t+1}^t$$  \hspace{1cm} (12)

where $Q_{1,n,t}^t$ denotes the discretised RND of term $t$ across the specified $n$ states in $P$. Equation (12) means that the RND of term $t + 1$ is equivalent to the product of the RND at term $t$ and the constant TPM. Tran and 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.$^9$

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

$$P = \arg\min_{P_{ij} \geq 0} \| AP - B \|_2^2.$$  \hspace{1cm} (13)

The $A$ and $B$ matrices can be quite large in practice though, 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.

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 using the estimation method given above, his replication was considerably different to Ross’ original results. This suggests that Ross includes additional constraints on the structure of the transition matrix. To this end, Spears (2013) tests nine alternative constrained estimation methods and, using a range of fitting criteria, finds 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.

Rather than impose constraints directly on the transition probabilities, Audrino et al. (2015) consider the alternative route of using Tikhonov regularization on the constrained

$^9$Although not shown below, we test several alternative state space grids in our algorithm below and find little difference in the recovered results.
OLS problem (Tikhonov and Arsenin (1977)). In essence, the idea of regularization 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. Classical Tikhonov regularization 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.

In this vein, Kiriu and 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 and Hibiki thus attempt to solve the following regularized OLS problem:

$$P = \arg\min_{p_{ij} \geq 0} \|AP - B\|_2^2 + \zeta \|P - \bar{P}\|_2^2$$  (14)

where $\zeta > 0$ is the regularization parameter that governs the weight given to the additional regularization norm. Setting $\zeta = 0$ returns the original OLS problem.

Although regularization 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 regularization value, $\zeta^*$. 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 and Hibiki (2015)). After testing each of the respective methods proposed in the above papers, we choose to adopt the $h$ function proposed by Kiriu and Hibiki (2015) for its appreciably better robustness. See Appendix B for function details.

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

1. Estimate standardised RNDs as per the procedure given in Section 2.3 and construct a discrete 21-state $Q$ matrix spanning a $50 - 150\%$ range of the prevailing spot level in 5% intervals.\(^{10}\)

2. 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 $B = Q'_{1:21,4:T}$.

3. Construct Kiriu & Hibiki’s (2015) regularization target matrix $\bar{P}$ and solve the regularized OLS problem in Equation 14 using the constrained linear least-squares solver in MATLAB for a wide range of regularization values, $\zeta = \left[0, 10^{-5}, 10^{-4.8}, \ldots, 10^2\right]$.

\(^{10}\)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 regularized OLS problems over the full data sample.
4. Find the regularization value that minimises the selection function, $\zeta^* = \arg\min h(\zeta)$.

5. Using the optimal $\zeta^*$, solve another regularized OLS problem on a finer 51-state $Q$-matrix in order to estimate a final risk-neutral transition probability matrix $P^*$.

6. 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 extract the discrete three month real-world return distribution as the middle row of the $F$ matrix.

Figure 1: Representative diagram of the real-world probability recovery process for options data from 23 Apr 2007.

Figure 2 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.
4 Top40 Option-Implied Distributions

We use the estimation algorithm outlined in Section 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 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.

Figure 2: Weekly box plot of three month risk-neutral Top40 distributions, Sep 2005 - Jan 2016

Before considering the implied distribution moments though, Figure 2 provides some intuition on how the Top40 three month RND changes over the full test period. The evolution of the complete distribution as well as individual percentiles are displayed here. This information can be used descriptively or prescriptively. In the descriptive sense, the negative tail of the distributions (red shading) is consistently much 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 considerably during the global financial crisis and remained wider than usual until late into 2009. Since then, the RND has narrowed considerably, although the negative tail has once again started to increase over the last couple of years. In the prescriptive sense, one could for example focus on the 95th 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.
4.1 Risk-Neutral and Real-World Implied Moments

We now consider the implied moments of the risk-neutral and recovered three month distributions. Figure 3 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 gives 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 actual annualised Top40 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.

From the second panel in Figure 3, we 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 results suggests that our recovery algorithm is giving us results that at least match the arbitrage pricing criteria.

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.

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

The lower right quadrant reveals the strong positive relationship between risk-neutral and real-world volatility as well as significantly positive correlations between the two skew-
Figure 3: Top40 real-world vs. risk-neutral moments graphed against index performance
ness 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.

Table 1: Implied moment summary statistics

|                      | Mean | Vol | Min 5% | 25th | 50th | 75th | 95th | Max  |
|----------------------|------|-----|--------|------|------|------|------|------|
| **Top40 Returns**1   | 16.1%| 21.5%| -11.7% | -4.7%| -1.2%| 0.5% | 2.1% | 4.6% | 15.3%|
| **Risk-Neutral**2    | 4.0% | 1.9% | 1.3%   | 1.8% | 2.4% | 3.3% | 4.9% | 7.9% | 9.0% |
| **Risk-Neutral**2    | 23.5%| 6.8% | 12.4%  | 15.2%| 18.8%| 22.1%| 26.6%| 34.7%| 55.8%|
| **Risk-Neutral**2    | -0.91| 0.29 | -1.67  | -1.398| -1.12 | -0.92 | -0.74 | -0.45 | 0.09 |
| **Risk-Neutral**2    | 4.95 | 1.24 | 2.68   | 3.20 | 4.16 | 4.75 | 5.62 | 7.21 | 10.97|
| **Risk-Neutral**2    | 15.1%| 3.6% | 0.6%   | 9.1% | 13.1%| 14.7%| 17.4%| 21.2%| 28.9%|
| **Risk-Neutral**2    | 21.3%| 6.6% | 10.3%  | 13.1%| 17.0%| 19.9%| 24.2%| 32.1%| 48.3%|
| **Risk-Neutral**2    | -1.05| 0.28 | -1.71  | -1.48| -1.25 | -1.07 | -0.87 | -0.62 | -0.08 |
| **Risk-Neutral**2    | 5.39 | 1.34 | 2.21   | 3.31 | 4.53 | 5.38 | 6.26 | 7.67 | 11.00|

1 Top40 Returns mean and volatility are annualised, percentiles are weekly numbers.
2 Mean and Volatility values are all annualised

Table 2: Correlation matrix of weekly implied moment changes

**4.2 Tactical Asset Allocation with Option-Implied Information**

We heuristically test the forward-looking information content of the moments by following a simple tactical asset allocation (TAA) strategy over the sample period 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. 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

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 brown shaded lines are the real-world strategies. Table 3 gives the summary statistics from the trading strategies.

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 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 re-

---

11 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.

12 Note that an investor will generally only want higher kurtosis and thus fatter tails when skewness is become increasingly positive.
covered real-world counterparts. We leave a proper discussion of this conjecture for later research though.

|                | Top40 | Risk-Neutral Moment TAA | Real-World Moment TAA |
|----------------|-------|-------------------------|-----------------------|
|                | Mean  | Volatility | Skewness | Kurtosis | Mean  | Volatility | Skewness | Kurtosis |
| Mean           | 16.12%| 14.25%      | 17.03%   | 16.67%   | 16.94%| 18.05%     | 17.51%   | 17.08%   |
| Volatility     | 21.71%| 17.14%      | 15.25%   | 16.66%   | 15.83%| 16.02%     | 15.84%   | 16.19%   |
| Sharpe Ratio   | 0.50  | 0.52        | 0.77     | 0.69     | 0.74  | 0.80       | 0.77     | 0.73     |
| Skewness       | -0.24 | -0.13       | -0.14    | 0.13     | 0.28  | 0.03       | 0.35     | 0.23     |
| Kurtosis       | 5.99  | 11.37       | 8.19     | 10.41    | 12.45 | 12.97      | 12.17    | 11.51    |
| No. Trades     | n.a.  | 289         | 243      | 283      | 296   | 269        | 293      | 291      |
| Max DD¹        | -46.17%| -41.54%    | -33.56%  | -26.78%  | -35.18%| -26.97%    | -38.04%  | -34.77%  |

Table 3: Top40 implied moment trading strategy results

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.

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

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.

We show that the deterministic SVI volatility model is a viable candidate for modelling implied volatility surfaces and use this model to estimate the underlying risk-neutral distri-
butional surfaces on the Top40 index. The issue of calibration with sparse and noisy data is considered at length and a simple but robust fitting algorithm is proposed.

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.
References

Aıt-Sahalia, Y. and J. Duarte (2003). Nonparametric Option Pricing under Shape Restrictions. *Journal of Econometrics* 116(1), 9–47.

Aït-Sahalia, Y. and A. W. Lo (1998). Nonparametric Estimation of State-price Densities implicit in Financial Asset Prices. *The Journal of Finance* 53(2), 499–547.

Audrino, F., R. Huitema, and M. Ludwig (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, D. S. (2000). Post-'87 Crash Fears in the S&P 500 Futures Option Market. *Journal of Econometrics* 94(1), 181–238.

Baule, R., O. Korn, and S. Saßning (2015). Which Beta Is Best? On the Information Content of Option-Implied Betas. *European Financial Management*.

Black, F. (1976). The Pricing of Commodity Contracts. *Journal of Financial Economics* 3(1), 167–179.

Black, F. and M. Scholes (1973). The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy*, 637–654.

Bliss, R. R. and N. Panigirtzoglou (2002). Testing the Stability of Implied Probability Density Functions. *Journal of Banking & Finance* 26(2), 381–422.

Bliss, R. R. and N. Panigirtzoglou (2004). Option-Implied Risk Aversion Estimates. *The Journal of Finance* 59(1), 407–446.

Borovička, Jaroslav and Hansen, Lars Peter and Scheinkman, José A (2016). Misspecified recovery. *The Journal of Finance*.

Breeden, D. T. and R. H. Litzenberger (1978). Prices of State-Contingent Claims implicit in Option Prices. *Journal of Business*, 621–651.

Buchen, P. W. and M. Kelly (1996). The Maximum Entropy Distribution of an Asset inferred from Option Prices. *Journal of Financial and Quantitative Analysis* 31(01), 143–159.

Buss, A. and G. Vilkov (2012). Measuring Equity Risk with Option-Implied Correlations. *Review of Financial Studies* 25(10), 3113–3140.
Carr, P., H. Geman, D. B. Madan, and M. Yor (2003). Stochastic Volatility for Lévy Processes. *Mathematical Finance* 13(3), 345–382.

Carr, P. and D. B. Madan (2005). A Note on Sufficient Conditions for No Arbitrage. *Finance Research Letters* 2(3), 125–130.

Carr, P. and J. Yu (2012). Risk, Return, and Ross Recovery. *Journal of Derivatives* 20(1), 38.

Cochrane, J. H. (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., E. Jondeau, and M. Rockinger (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, J. C. and S. A. Ross (1976). The Valuation of Options for Alternative Stochastic Processes. *Journal of Financial Economics* 3(1-2), 145–166.

DeMarco, P. and C. Martini (2009). Quasi-Explicit Calibration of Gatheral SVI model. *Zeliade Systems*.

DeMiguel, V., Y. Plyakha, R. Uppal, and G. Vilkov (2013). Improving Portfolio Selection Using Option-Implied Volatility and Skewness. *Journal of Quantitative & Financial Analysis* 48(6), 1813–1845.

Dubynskiy, S. and R. S. Goldstein (2013). Recovering Drifts and Preference Parameters from Financial Derivatives. *Available at SSRN 2244394*.

Dumas, B., J. Fleming, and R. E. Whaley (1998). Implied Volatility Functions: Empirical Tests. *The Journal of Finance* 53(6), 2059–2106.

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. and A. Jacquier (2014). Arbitrage-free SVI Volatility Surfaces. *Quantitative Finance* 14(1), 59–71.

Hagan, P. S., D. Kumar, A. S. Lesniewski, and D. E. Woodward (2002). Managing Smile Risk. *The Best of Wilmott*, 249.

Heston, S. L. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency options. *Review of Financial Studies* 6(2), 327–343.

Kempf, A., O. Korn, and S. Saßning (2015). Portfolio Optimization using Forward-looking Information. *Review of Finance* 19(1), 467–490.

Kiriu, T. and N. Hibiki (2015). Estimating Forward Looking Distribution with the Ross Recovery Theorem. *Keio University Working Paper*.

Kostakis, A., N. Panigirtzoglou, and G. Skiadopoulos (2011). Market Timing with Option-implied Distributions: A Forward-looking Approach. *Management Science* 57(7), 1231–1249.

Kotzé, A. and A. Joseph (2009). Constructing a South African Index Volatility Surface from Exchange Traded Data. *Available at SSRN 2198357*.

Malz, A. M. (2014). A Simple and Reliable Way to Compute Option-based Risk-neutral Distributions. Technical Report 677, FRB of New York.

Martin, I. and S. Ross (2013). The Long Bond. Technical report, London School of Economics and MIT.

McManus, D. J. et al. (1999). *The Information Content of Interest Rate Futures Options*. Citeseer.

Melick, W. R. and C. P. Thomas (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, R. C. (1973). Theory of Rational Option Pricing. *The Bell Journal of Economics and Management Science*, 141–183.

Merton, R. C. (1976). Option Pricing when Underlying Stock Returns are Discontinuous. *Journal of Financial Economics* 3(1-2), 125–144.

Posner, S. E. and M. A. Milevsky (1998). Valuing Exotic Options by Approximating the SPD with Higher Moments. *The Journal of Financial Engineering* 7(2).

Roper, M. (2010). Arbitrage Free Implied Volatility Surfaces. Technical report, University of Sydney Working Paper.
REFERENCES

Ross, S. (2015). The Recovery Theorem. *The Journal of Finance* 70(2), 615–648.

Ross, S. A. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory* 13(3), 341–360.

Shimko, D. (1993). The Bounds of Probability. *RISK* 6, 33–37.

Shreve, S. E. (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. and V. Y. Arsenin (1977). *Methods for Solving Ill-Posed Problems*. John Wiley and Sons, Inc.

Tompkins, R. G. (2001). Implied Volatility Surfaces: Uncovering Regularities for Options on Financial Futures. *The European Journal of Finance* 7(3), 198–230.

Tran, N.-K. and S. Xia (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), 371–385.
A

SVI Calibration Algorithm

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

\[
y = \frac{x - m}{s}.
\]  
(15)

Using this new variable we can rewrite the SVI equation in total variance space as

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

where

\[
\alpha = a \tau \\
\beta = b s \tau \\
\delta = \rho b s \tau.
\]  
(17)

Taking \( m \) and \( s \) as fixed values, the objective function then becomes

\[
\arg\min_{(\alpha, \beta, \delta) \in D} \sum_{i=1}^{n} w_i \left( \alpha + \delta y + \beta \sqrt{y^2 - 1} - V_i \right)^2,
\]  
(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

\[
D = \begin{cases}
0 \leq \beta \leq 4s \\
|\delta| \leq \beta \text{ and } |\delta| \leq 4s - \beta \\
0 \leq \alpha \leq \max \{v_i\}.
\end{cases}
\]  
(19)

For a given solution set \( \{\alpha^*, \beta^*, \delta^*\} \), one can find the corresponding solution set \( \{a^*, b^*, \rho^*\} \) and thus the remaining parameters can be calibrated as follows

\[
\arg\min_{m, s \geq 0} \sum_{i=1}^{n} w_i \left( \sigma^2(x, \{m, s, a^*, b^*, \rho^*\}) - v_i \right)^2.
\]  
(20)

The original 5-dimensional calibration is thus broken into separate 2-dimensional and three-dimensional minimisation problems.

B

Regularization Parameters

Kiriu and Hibiki (2015) calculate the regularisation target matrix \( \bar{P} \) directly from the discretised RND matrix \( Q_{1:n, 1:T-1}^t \) 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 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 $k \leq \min{(n-i,n-j)}$. The first premise defines the middle row of $\bar{P}$ while the second defines the remainder of the matrix:

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

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

$$P = \arg\min_{P \geq 0} \|AP - B\|_2^2 + \zeta \|P - \bar{P}\|_2^2 = \arg\min_{P \geq 0} y_{\text{fit}}(\zeta) + \zeta y_{\text{reg}}(\zeta).$$ (22)

The selection function is then given as

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

where the respective denominators represent the maximum spread in each term and the numerator gives the spread achieved for a specified $\zeta$ value. The values $y(0)$ are solutions from the original OLS problem and $y_{\text{reg}}(\infty)$ is set to 0 due to the fact that $P \to \bar{P}$ as $\zeta \to \infty$. This then means that $h(0) = h(\infty) = 1$. Kiriu and Hibiki (2015) further 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 which makes it a very appealing selection function.