Estimating Forward Looking Distribution with the Ross Recovery Theorem

Takuya Kiriu†
Graduate School of Science and Technology
Keio University, Yokohama, Japan
Tel: (+81) 45-566-1454, Email:kiriutakuya@gmail.com

Norio Hibiki
Faculty of Science and Technology
Keio University, Yokohama, Japan
Tel: (+81) 45-566-1635, Email: hibiki@ae.keio.ac.jp

Abstract. The payoff of option is determined by the future price of underlying asset and therefore the option prices contain the forward looking information. Implied distribution is a forward looking distribution of the underlying asset derived from option prices. There are a lot of studies estimating implied distribution in the risk neutral probability framework. However, a risk neutral probability generally differs from a real world probability, which represents actual investors view about asset return. Recently, Ross (2015) has showed remarkable theorem, named "Recovery Theorem". It enables us to estimate the real world probability distribution from option prices under a particular assumption about representative investor's risk preferences. However, it is not easy to derive the appropriate estimators because it is necessary to solve an ill-posed problem in estimation process. This paper discusses about the method to estimate a real world distribution accurately with the Recovery Theorem. The previous studies propose the methods to estimate the real world distribution, whereas they do not investigate on the estimation accuracy. Hence, we test the effectiveness of the Tikhonov method used by Audrino *et al.* (2015) in the numerical analysis with hypothetical data. We propose a new method to derive the more accurate solution by configuring the regularization term considering prior information and compare it with the Tikhonov method. Moreover, we discuss regularization parameter selection to get the accurate real world distribution. We find the following three points through the numerical analysis.

(1) To stabilize the solution by introducing regularization term is an effective method in terms of estimating a real world distribution with the Recovery Theorem.
(2) Proposed method can estimate a real world distribution more accurately than the Tikhonov method.
(3) We can offer the appropriate solutions even if the number of maturities is less than that of states.

Keywords: finance, implied distribution, Recovery Theorem, regularization, prior information

1. INTRODUCTION

The payoff of option is determined by the future price of underlying asset and therefore the option prices contain the forward looking information. Implied distribution is a forward looking distribution of the underlying asset derived from option prices, which is useful for decision making in financial market such as development of investment strategy and monetary policy. It is possible to derive a risk neutral distribution from option prices in a complete market, and there are a lot of studies on the distribution. However, the risk neutral probability is generally different from the real world probability, and the real world distribution expresses actual investor's view.

Recently, Ross (2015) has showed remarkable theorem, named "Recovery Theorem". It enables us to estimate a real world distribution from option prices under a particular assumption about representative investor's risk preferences. There are two types of studies related to the Recovery Theorem. The first is the theoretical extension into the continuous time case (See Carr and Yu (2012), Dubynskiy and Goldstein (2013), Walden (2014), Park (2015) and Qin and Linetsky (2015) and the fixed income market (Martin and Ross (2013)). The second is the development of the practical methodology to estimate real world distribution from option prices. Spears (2013) indicates that estimators derived by the simple and instructive method of Ross (2015) are intuitively inaccurate, and compares the estimators under various constraints. Audrino *et al.* (2015) point out that it is necessary to solve an ill-posed problem in estimation process, and propose to apply Tikhonov method, which is a standard regularization method for ill-posed problems. In addition, they estimate a real world distribution from 13 years of S&P500 option data and investigate the effectiveness of simple investment strategy based on moments of the distribution. To the best of our knowledge, this is the only research that uses time series data. Backwell (2015) denotes time-homogeneity of state prices, which is hypothesized when estimating real world distribution, cannot be realized in a real market. The estimation method is also proposed to reduce the bias.

Our paper is included in the second type and discusses about the method to estimate a real world distribution accurately with the Recovery Theorem. The previous studies propose the methods to estimate the real world distribution, whereas they do not investigate the estimation accuracy. It is important to examine the problem because it is ill-posed in estimation process. Hence, we test the effectiveness of the Tikhonov method, in the numerical analysis with hypothetical data. We propose a new method to derive the more accurate solution by configuring the regularization term reconsidering prior information, and compare it with the Tikhonov method. Moreover, we discuss regularization parameter selection to get the accurate real world distribution.

We find the following three points through the numerical analysis. (1) To stabilize the solution by introducing regularization term is an effective method in terms of estimating a real world distribution with the Recovery Theorem. (2) Proposed method can estimate a real world distribution more accurately than Tikhonov method. (3) We can offer the appropriate solutions even if the number of maturities is less than that of states.

2. RECOVERY THEOREM

In this section, we summarize the Recovery Theorem. We assume an arbitrage free and complete market in discrete time with finite state one period model. Market states $\theta_i \ (i = 1, \dots, n)$ are defined by r_i, which are underlying stock index returns from time 0. $P := (p_{i,j})$ is a $n \times n$ transition state price matrix. $p_{i,j}$ is a state price from θ_i to θ_j.[1] We similarly define a $n \times n$ risk neutral transition probability matrix $Q := (q_{i,j})$ and a $n \times n$ real world transition probability matrix $F := (f_{i,j})$. We also describe the notation Q as "risk neutral distribution" and F as "real world distribution" depending on the context. P is assumed to be irreducible[2], and therefore Q and F are also irreducible.

In this section, we suppose that P is known because it can be estimated from option prices.[3] Q is easily derived from P, since $q_{i,j}$ is expressed as follows,

$$q_{i,j} = \frac{p_{i,j}}{\sum_{k=1}^{n} p_{i,k}} \quad (i, j = 1, \dots, n). \tag{1}$$

On the other hand, it is difficult to derive F because the state price is simultaneously a function of both a real world probability and market risk preferences. However, Ross (2015) showed F can be derived from P under the assumption that there is a representative investor with Time Additive Intertemporal Expected Utility Theory preferences over consumption (TAIEUT investor). A utility function of the TAIEUT investor is given by

$$U(c_i) + \delta \sum_{j=1}^{n} f_{i,j} U(c_j) \quad (i = 1, \dots, n), \tag{2}$$

where c_i is the consumption at θ_i, $U(c)$ is a utility for the consumption and $\delta\ (> 0)$ is the discount factor of the utility. We assume that $U(c)$ holds nonsatiation condition $U'(c) > 0$ but do not restrict its parametric form. Then the relationship between $f_{i,j}$ and $p_{i,j}$ is expressed as

$$f_{i,j} = \frac{1}{\delta} \frac{U'(c_i)}{U'(c_j)} p_{i,j} \quad (i, j = 1, \dots, n). \tag{3}$$

The ratio of $p_{i,j}$ to $f_{i,j}$ is called pricing kernel, and it is expressed as

$$\phi_{i,j} := \frac{p_{i,j}}{f_{i,j}} = \delta \frac{U'(c_j)}{U'(c_i)} \quad (i, j = 1, \dots, n). \tag{4}$$

It is dependent on investor's risk preferences.

Since P is non-negative and irreducible, the Perron-Frobenius Theorem asserts that P has a unique strictly positive eigenvector $\boldsymbol{v}$ associated with the maximum eigenvalue λ. The Recovery Theorem says that $\delta = \lambda$ and $U'(c_i) = v_i^{-1}\ (i = 1, \dots, n)$ hold, where v_i denotes the $i-$th element of $\boldsymbol{v}$.

---

[1] The state price $p_{i,j}$ shows the price of the security at θ_i which pays one dollar if the next state becomes θ_j and nothing otherwise.

[2] Irreducibility is defined as existing $k \in \mathbb{N}$ which satisfies $(P^k)_{i,j} > 0$ for all i, j. This assumption is very likely to be held.

[3] This is explained in Section 3 in detail.

We can calculate F from P with the Recovery Theorem as follows. We solve the eigenvalue problem of P and derive the maximum eigenvalue λ and the corresponding eigenvector $\boldsymbol{v}$. Then, we can calculate the elements of the matrix F as

$$f_{i,j} = \frac{1}{\lambda}\frac{v_j}{v_i}p_{i,j} \quad (i,j = 1,\dots,n). \tag{5}$$

In addition, Ross (2015) also proves that the real world distribution becomes equal to the risk neutral distribution, or $F = Q$, when the sum of the row elements of P is the same for each row, and it is a special case of the Recovery Theorem.

3. IMPLEMENTATION

We assume that P is known in Section 2, however it is necessary to estimate P from market option prices in practice. We represents the estimation procedure as referred to Spears (2013) in Figure 1. This section discusses Step 1 and Step 2 because the Recovery Theorem is simply applied in Step 3. Moreover, we point out the problem which occurs in Step 2 and propose a new method.

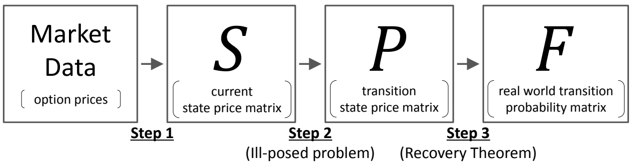


Figure 1: Process of recovery

3.1 Step 1: From Option Prices to S

A $n \times m$ current state price matrix is defined as $S := (s_{j,\tau})$, where $s_{j,\tau}$ is a current state price for $\tau\ (= 1,\dots,m)$ periods transition from current state θ_{i_0} to θ_j. For simplicity, we assume the number of states is odd and θ_{i_0} is the center state $(i_0 = (n+1)/2)$.

We estimate S from option prices in Step 1. A method proposed by Breeden and Litzenberger (1978) is often used to estimate S and it is used to calculate the state price more accurately in a lot of literatures. It is not difficult to estimate S, and therefore we focus on Steps 2 and 3 in the analysis.

3.2 Step 2: From S to P

In Step 2, we estimate the $n \times n$ matrix P from the $n \times m$ matrix S assuming that transitions of the states follow time-homogeneous Markov chain.

We assume it is satisfied that $n \geq m$, except the analysis in Section 4.5. Denote the first column vector of S by $\boldsymbol{s}_1$, and the i_0-th row vector of P by $\boldsymbol{p}_{i_0}$. The j-th element of both vectors are $p_{i_0,j}$ according to the definition. Namely,

$$\boldsymbol{s}_1 = \boldsymbol{p}_{i_0}. \tag{6}$$

Because P represents the state transition in one period, we have the following relationship among $\boldsymbol{s}_\tau, \boldsymbol{s}_{\tau+1}$ and P.

$$\boldsymbol{s}_{\tau+1}^\top = \boldsymbol{s}_\tau^\top P \ (\tau = 1,\dots,m-1) \tag{7}$$

Denote the $(m-1)\times n$ matrix transposed from S except the last column by A, and the $(m-1)\times n$ matrix transposed from S except the first column by B. Equation (7) can be expressed as follows.

$$AP = B \tag{8}$$

P should be estimated by minimizing the differences in both sides of Equation (8) under the no-arbitrage conditions and Equation (6). The mathematical formulation is

$$\min_P \quad \|AP - B\|_2^2 \tag{9}$$

$$\text{subject to} \quad \boldsymbol{s}_1 = \boldsymbol{p}_{i_0} \tag{10}$$

$$p_{i,j} \geq 0 \ \ (i,j = 1,\dots,n). \tag{11}$$

Audrino *et al.* (2015) indicate that the average condition number of 11×11 matrix A estimated from S&P 500 option data from 2000 to 2012 is a very large value of 2.17×10^8, and therefore the problem is ill-posed. The ill-posed problem has a set of candidates of optimal solutions whose objective function values are almost the same due to low independency of data. Consequently, it has the bad characteristics that the solution is highly sensitive to a small noise. Then, Audrino *et al.* (2015) propose to use the Tikhonov method, which is a standard regularization method, in order to solve the ill-posed problem. The regularization method is formulated by adding the regularization term to the objective function to stabilize the solution against a small change of the input parameter. The regularization term gives the prior information about the expected characteristics of solution. Specifically, the objective function is reformulated as follows,

$$\min_P \quad \|AP - B\|_2^2 + \zeta\|P\|_2^2. \tag{12}$$

The second term is a regularization term and $\|\cdot\|_2$ denotes the Euclidean norm. ζ is called a regularization parameter and controls the trade-off between fitting and stability. Equation (12) can be transformed using a $n \times n$ unit matrix I and a null matrix O.

$$\min_P \quad \left\| \begin{bmatrix} A \\ \sqrt{\zeta} I \end{bmatrix} P - \begin{bmatrix} B \\ O \end{bmatrix} \right\|_2^2 \tag{13}$$

In the Tikhonov method, the problem is solved with the prior information that the small $p_{i,j}$ is preferable. However, it is inadequate to estimate P using the information because $p_{i,j}$ should be larger for the higher transition probability. In addition, the matrix P is not irreducible in the special case of $\zeta \to \infty$. This means that the Recovery Theorem cannot be always applied for any ζ. Therefore, it is difficult to interpret the relationship between ζ and the real world distribution F.

We configure the regularization term for the two preferable prior information to estimate P as follows.

1. $\boldsymbol{s}_1$ is equal to $\boldsymbol{p}_{i_0}$ (Equation (6)). It is theoretically derived as above-mentioned.
2. $p_{i,j}$ is similar to $p_{i+k,j+k}$ $(i,j = 1,\dots,n; k \in \mathbb{Z},\ 1 \le i+k \le n, 1 \le j+k \le n)$. This means the state prices with the equal difference of transition between states are similar to each other. It is not the theoretically-derived condition, but it is empirically expected.

We propose a new method so that we can configure the regularization term for the prior information mentioned above. Specifically, we rewrite Equation (9) into

$$\min_P \quad \|AP - B\|_2^2 + \zeta\|P - \bar{P}\|_2^2 \tag{14}$$

$$\Leftrightarrow \min_P \quad \left\| \begin{bmatrix} A \\ \sqrt{\zeta} I \end{bmatrix} P - \begin{bmatrix} B \\ \sqrt{\zeta}\bar{P} \end{bmatrix} \right\|_2^2 \tag{15}$$

where,

$$\bar{P} = \begin{bmatrix} \bar{p}_{1,1} & \bar{p}_{1,2} & \cdots & \bar{p}_{1,i_0} & \cdots & \bar{p}_{1,n-1} & \bar{p}_{1,n} \\ \bar{p}_{2,1} & \bar{p}_{2,2} & \cdots & \bar{p}_{2,i_0} & \cdots & \bar{p}_{2,n-1} & \bar{p}_{2,n} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots \\ \bar{p}_{i_0,1} & \bar{p}_{i_0,2} & \cdots & \bar{p}_{i_0,i_0} & \cdots & \bar{p}_{i_0,n-1} & \bar{p}_{i_0,n} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots \\ \bar{p}_{n-1,1} & \bar{p}_{n-1,2} & \cdots & \bar{p}_{n-1,i_0} & \cdots & \bar{p}_{n-1,n-1} & \bar{p}_{n-1,n} \\ \bar{p}_{n,1} & \bar{p}_{n,2} & \cdots & \bar{p}_{n,i_0} & \cdots & \bar{p}_{n,n-1} & \bar{p}_{n,n} \end{bmatrix} \tag{16}$$

$$= \begin{bmatrix} \sum_{k=1}^{i_0} s_{k,1} & s_{i_0+1,1} & \cdots & s_{n,1} & \cdots & 0 & 0 \\ \sum_{k=1}^{i_0-1} s_{k,1} & s_{i_0,1} & \cdots & s_{n-1,1} & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots \\ s_{1,1} & s_{2,1} & \cdots & s_{i_0,1} & \cdots & s_{n-1,1} & s_{n,1} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & s_{2,1} & \cdots & s_{i_0,1} & \sum_{k=i_0+1}^{n} s_{k,1} \\ 0 & 0 & \cdots & s_{1,1} & \cdots & s_{i_0-1,1} & \sum_{k=i_0}^{n} s_{k,1} \end{bmatrix}. \tag{17}$$

Our method stabilizes the elements of the estimated matrix P by getting closer to $\bar{P}$, which expresses the prior information. However, the values are accumulated in the first and last columns of the matrix, and we set zero to the other elements. We have the same sensitivity of P to small change of input value for the Tikhonov method and our method because the first term of Equation (13) is the same matrix as that of Equation (15). Our method can clarify the effects of the regularization term on the real world distribution F. The sum of the elements for every row of a matrix $\bar{P}$ is identical. In the case of $\zeta \to \infty$, we obtain $P = \bar{P}$, and the real world probability coincides with risk neutral probability. Therefore, as ζ gets larger, it is expected that the estimated matrix F gets closer to Q. Our method can also derive a risk neutral distribution as forward looking distribution in the framework of the Recovery Theorem.

4. NUMERICAL ANALYSIS

We verify the accuracy of estimation and examine the effectiveness of the proposed method. However, it is difficult to know a true real world distribution from real data. Therefore we assume real world distribution using hypothetical data and verify the accuracy of the estimates.

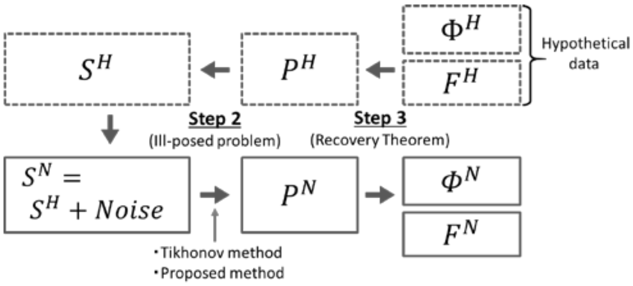


Figure 2: Framework of analysis

Figure 2 represents framework of the analysis. Firstly, we give the two hypothetical matrices; hypothetical real world transition probability matrix F^H and pricing kernel matrix Φ^H. Then, we calculate the transition state price matrix P^H and current state price matrix S^H in backward order. In reality, it is difficult to obtain S^H because of the noise contained option price data and estimation error of Step 1, so we generate the matrix S^N given by adding noise to S^H. We assume the noise $e_{i,j}$ follows i.i.d., and normal distribution with mean 0 and standard deviation σ.

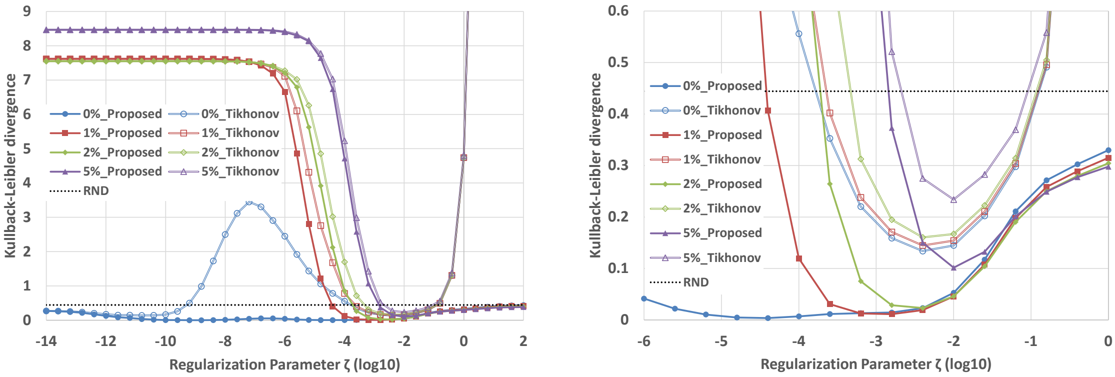
$$s_{i,j}^N = s_{i,j}^H(1 + e_{i,j}) \quad (i,j = 1,\dots,n) \tag{18}$$

In this way, we eliminate the impact of estimation method of Step 1. Then, we estimates P^N from S^N (Step 2) with proposed or Tikhonov method, and derive F^N and Φ^N applying the Recovery Theorem to P^N (Step3). If the estimator F^N is close to the original data F^H, the estimation method is appropriate because we can restore the original data. The detail of estimation accuracy criteria is mentioned in Section 4.1.

4.1 Setting

We explain the definition of states, the way of generating hypothetical data, and the evaluation criteria of estimation accuracy.

Market state is defined by return from time 0. We provide 31 returns placed by 2% symmetrically from the return of 0%, which is equally divided from −30% to 30% and i_0=16.



[a] General view [b] Enlarged view

Figure 3: Base case: KL divergence of $\boldsymbol{f}_{i_0}^N$ from $\boldsymbol{f}_{i_0}^H$

We apply the number of maturities of option traded in the market to the number of period m when estimating S from real data. However, we can apply any value of m for the hypothetical data. We analyze the case where the number of estimated variables is the same as the number of data, or $m = n = 31$. The case where $m < n$ is analyzed in Section 4.5.

We denote the i_0-th row vector of the matrix F by $\boldsymbol{f}_{i_0}$, which is the real world distribution at current state. We evaluate the estimation accuracy by the KL divergence of $\boldsymbol{f}_{i_0}^N$ from $\boldsymbol{f}_{i_0}^H$. It is a standard measure of the difference between the two distributions and defined as

$$D_{KL}\left(\boldsymbol{f}_{i_0}^N \middle| \boldsymbol{f}_{i_0}^H\right) := \sum_{j=1}^{n} f_{i_0,j}^N \ln\left(\frac{f_{i_0,j}^N}{f_{i_0,j}^H}\right). \quad (19)$$

When the two distributions are exactly equal, KL divergence is equal to zero. We also have the same conclusion in the cases of evaluating entire matrix and using a different criteria such as Euclidean distance. Hence, we show only the result using $D_{KL}\left(\boldsymbol{f}_{i_0}^N \middle| \boldsymbol{f}_{i_0}^H\right)$ hereafter.

4.2 Hypothetical Data

We give the hypothetical matrix Φ^H and F^H as plausible as possible used in the analysis.

The matrix Φ^H is generated by assuming TAIEUT investor who has a CRRA utility function $U(c) = c^{1-\gamma_R}/(1-\gamma_R)$. γ_R is relative risk aversion. Assuming TAIEUT investor, ϕ can be decomposed into U and δ shown in Equation (4). So, we can denote the (i,j) element of Φ^H by

$$\phi_{i,j}^H = \delta\left(\frac{1+r_j}{1+r_i}\right)^{-\gamma_R} \quad (i,j = 1,\ldots,n). \quad (20)$$

$\gamma_R = 3$ and $\delta = 0.999$ are used in the base case.

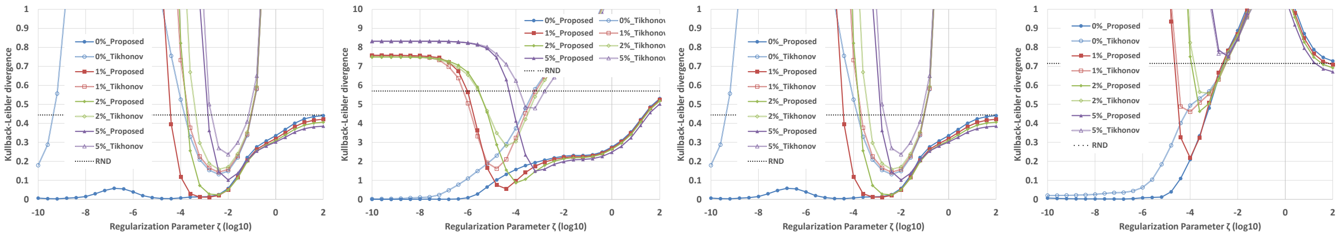
The matrix F^H is generated from the S&P 500 historical data. We set a reference date, and calculate returns from the reference date to the twelve dates which are set as every 30 calendar days. If it is a holiday, the day before a holiday is used. A matrix is generated by counting the number of state transitions in one period from the return sequence. Denote the return of state θ_j by r_j in the matrix, which is discretely described by every 2%. When a real historical return is between $r_j - 1\%$ and $r_j + 1\%$, it is assigned to state θ_j. A return more than or equal to 29% (less than or equal to -29%) are assigned to 30% (-30%). This is repeated daily by changing the reference date from Jan 3, 1950 to Jan 3, 2014. Then, all the matrices are summed up. Finally, each element of summed matrix is divided by each sum of the row elements to make it probability matrix.

The optimization problem in Step 2 is still ill-posed because the condition number of A^H calculated backward using Φ^H and F^H is very large, and 3.3×10^{16}. The numerical results are calculated using random noises for the specific random seed, but we derive the similar conclusions for the different seeds.

4.3 Base Case

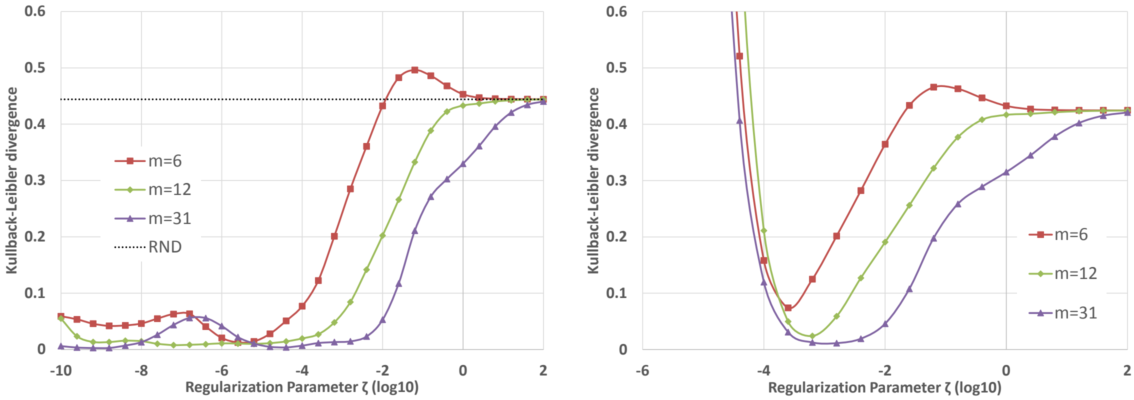
Figure 3 displays the KL divergence with the proposed method and Tikhonov method for different values of ζ where $\zeta = 10^{-14}, 10^{-13.6}, \ldots, 10^{1.6}, 10^{2}$. The KL divergence of $\boldsymbol{q}_{i_0}^H$ from $\boldsymbol{f}_{i_0}^H$, $D_{KL}\left(\boldsymbol{q}_{i_0}^H \middle| \boldsymbol{f}_{i_0}^H\right)$, is shown as "RND" (Risk Neutral Distribution). $\boldsymbol{q}_{i_0}^H$ is the most accurate distribution when we estimate the forward looking distribution in risk neutral probability framework. Getting a smaller KL divergence than RND is one of the important points to judge that a real world distribution estimator is good.

Firstly, we discuss about the results of the case of $\sigma = 0\%$ where S^N is observed without noise. Theoretically, the KL divergence where $\zeta = 0$ becomes zero because the distribution estimated without using regularization methods equals the original distribution. However, the estimated KL divergence is 0.273 due to the



[a] $\delta = 0.995$ [b] $\gamma_R = 10$ [c] U: CARA utility [d] F^{True}: Nikkei225

Figure 4: Robustness check: KL divergence of $\boldsymbol{f}_{i_0}^N$ from $\boldsymbol{f}_{i_0}^H$



[a] $\sigma = 0\%$ [b] $\sigma = 1\%$

Figure 5: The case where $m < n$: KL divergence of $\boldsymbol{f}_{i_0}^N$ from $\boldsymbol{f}_{i_0}^H$

calculation error. This shows that how it is difficult to get an accurate estimator of ill-posed problem. Using the proposed method, the minimum KL divergence is 2.65×10^{-3} where $\zeta = 10^{-9.2}$ and the estimation accuracy is improved drastically. In addition, the proposed method stabilize the estimators by achieving them closer to RND as ζ gets larger. Using the Tikhonov method, the KL divergence is also improved to 0.134 where $\zeta = 10^{-2.4}$. However, the KL divergence with proposed method is always lower than that with the Tikhonov method in any ζ.

We check the cases with noise ($\sigma = 1\%, 2\%, 5\%$). For small ζ, the estimation accuracy significantly deteriorate, compared with no noise case ($\sigma = 0\%$), because the problem is still ill-posed. The KL divergences decrease in both regularization methods, as ζ is greater to some extent. The result indicates that it is effective to introduce the regularization term in order to stabilize the solution, and the estimation accuracy in proposed method is better than Tikhonov method for any ζ. We find the low bias estimator is derived by the proposed method because the regularization term is configured more appropriately using the prior information.

4.4 Robustness Check

We check the robustness of the result in the base case by using the different hypothetical data from the base case. Figure 4a indicates the relationship between the regularization parameter ζ and KL divergence where we use $\delta = 0.995$. Figure 4b shows the result of $\gamma_R = 10$. Figure 4c displays the case of the CARA utility function with $\gamma_A = 3$ and Figure 4d shows the result of F^H estimated from Nikkei225 historical data of the same period in place of S&P500. These results show that the δ and utility function type are not sensitive to the KL divergence, but γ_R and F^H are sensitive to the shapes of graph. The following two features observed in the base case are preserved in any case. Firstly, the estimators using the proposed method or the Tikhonov method are more accurate than the estimators without regularization. Secondly, the estimation accuracy of proposed method is better than the Tikhonov method. The impact of hypothetical data change is not so big, and it is expected to get the similar results in most cases, as long as we use the plausible hypothetical data. However, the further analysis is required to demonstrate the robustness.

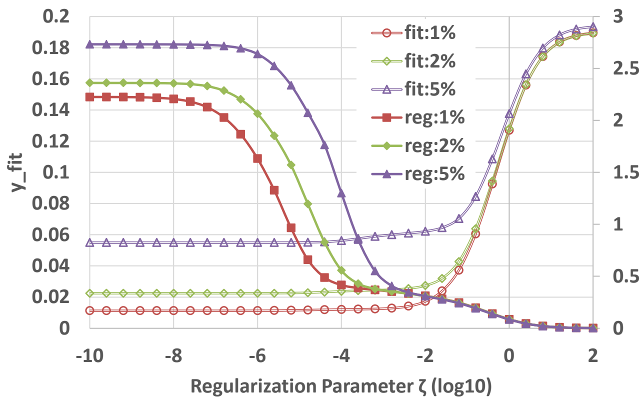


Figure 6: Decomposition of the objective function value

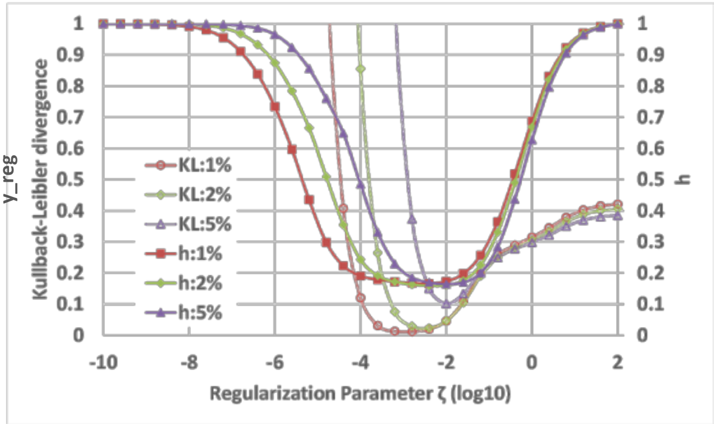


Figure 7: Function value of $h(\zeta)$ and KL divergence

4.5 The Case where $m < n$

The analysis is done with $m = 31$ so far. However, the number of option maturities traded typically in the market is less than 31. For instance, the number of S&P500 option maturities traded every month regularly in Chicago Board Options Exchanges is 12, and the number of Nikkei 225 option in Osaka Exchange is 9. In addition, m of S estimated from market data will be smaller because long-term options are likely to have low liquidity. We conduct the analysis for the case where the number of maturities (data) m is less than the number of states (estimated variables) n.[4] More specifically, we estimate the real world distribution under the 31 states and three kinds of the numbers of column of S $(m = 6, 12, 31)$ by the proposed method, and calculate their KL divergences.

Figure 5a shows the result of $\sigma = 0\%$. Even in the cases of $m = 6$ and 12, the variables can be estimated as accurately as the case of $m = 31$. The similar result is obtained in the case of $\sigma = 1\%$ in Figure 5b as well. It might be considered that it is impossible to get the accurate estimators since the number of data is less than the number of estimated variables. However, we can estimate the accurate estimators using the proposed method. This is because the prior information included in the regularization term offsets the insufficient information. In other words, all necessary information concerning investor's risk preferences is almost included in the matrix of six columns to estimate the real world distribution from the state price matrix.

[4] Usually, n should be less than m to estimate variables under the sufficient data. Therefore, the case in this section is analyzed under the insufficient uncertainty.

4.6 Selecting Regularization Parameter

We evaluate the estimation accuracy for various regularization parameters ζ, using the KL divergence. However, it is difficult to choose an appropriate value of ζ as a practical matter. Then, we propose a method of how to select ζ to get the accurate estimates of the real world distribution.

The objective function of optimization problem in Step 2 is Equation (14), and it consists of two parts. The first term shows the fitting error, and denote it by y_{fit}, whereas the second term except ζ shows the deviation between P^N and $\overline{P^N}$, and denote it by y_{reg}.We show them for the various ζ in Figure 6. As ζ increases, y_{fit} decreases and y_{reg} increases monotonically. Both y_{fit} and y_{reg} have the domain where the values greatly change. For example, in the case of $\sigma = 1\%$, the value of y_{fit} greatly changes around $\zeta = 10^{-6}$ and y_{reg} around $\zeta = 1$. This is one of the characteristics of the ill-posed problem. The purpose of using regularization term in the ill-posed problem is to find the optimal solution stably among the degenerated solutions which have almost the same fitting error, based on the prior information. Therefore, the sound strategy is to select ζ in the ranges where both y_{fit} and y_{reg} are relatively small. In the case of Figure 6, the appropriate value of ζ is between about 10^{-4} and 10^{-2}. In consideration of the fact that the range of y_{fit} is different from the range of y_{reg}, we propose the method of selecting ζ by minimizing the function $h(\zeta)$ defined as,

$$h(\zeta) := \frac{y_{fit}(\zeta) - y_{fit}(0)}{y_{fit}(\infty) - y_{fit}(0)} + \frac{y_{reg}(\zeta) - y_{reg}(\infty)}{y_{reg}(0) - y_{reg}(\infty)} \qquad (21)$$

$y_{fit}(\zeta)$ and $y_{reg}(\zeta)$ are functions of ζ as shown in Figure 6. $h(\zeta)$ is the sum of the normalized values. $y(0)$ is the value without the regularization term and $y(\infty)$ is the value derived under the condition that $P^N = \overline{P^N}$. So, $y_{reg}(\infty) = 0$ must hold. Moreover, $h(0) = 1$ and

$h(\infty) = 1$ must hold because both $y_{fit}(\zeta)$ and $y_{reg}(\zeta)$ are monotonic functions. We obtain the different values of $h(\zeta)$ by solving the optimization problems for the different values of ζ, and then we can adopt the ζ that minimizes $h(\zeta)$.

We show the function values of $h(\zeta)$ and the KL divergences for various values of ζ in the base case in Figure 7. In the range of small $h(\zeta)$, the KL divergence is also small in each case where $\sigma = 1\%, 2\%, 5\%$. It indicates that ζ can be selected well by using $h(\zeta)$. Our selection method could select appropriate values of ζ in most cases even for different hypothetical data, and we can find it effectively.[5]

5. CONCLUSION

The Recovery Theorem makes it possible to estimate the real world distribution implied in option prices. However, it is not easy to find accurate estimators because it is necessary to solve the ill-posed problem in the estimation process. This paper proposes the method to estimate the real world distribution accurately, and then analyzes how accurate the estimation is by numerical analysis using hypothetical data.

We clarify the following three points through the analysis. First, the regularization method like Tikhonov or our method used in Step 2 improves the estimation accuracy. This is because the regularization term enables us to suppress the effect of perturbation such as numerical error and data noise. Second, our method can estimate the real world distribution more accurately than the Tikhonov method, because our method could introduce more adequate regularization term, based on the prior information. Last, we find the fact that we derive the estimators accurately by our method to some extent even if the number of maturities of option is less than the number of states. It is sufficient to provide the six maturities of options in order to solve the problem with 31 states appropriately. This is likely to be less than the number of maturities of option traded in the market. The result suggests the possibility of obtaining the good estimator of the real world distribution from option prices traded in the market.

Future works are as follows, (1) checking the robustness of the result under more various conditions and hypothetical data, and (2) estimating a forward looking real world distribution from time-series option data with our proposed method and testing the predictability.

[5] We may need to compare it with alternative methods. For instance, the following function is considered,

$$h(\zeta) := \max\left(\frac{y_{fit}(\zeta) - y_{fit}(0)}{y_{fit}(\infty) - y_{fit}(0)}, \frac{y_{reg}(\zeta) - y_{reg}(\infty)}{y_{reg}(0) - y_{reg}(\infty)}\right) \quad (22)$$

However, we could select better ζ slightly using Equation (21), rather than Equation (22). Comparing with other alternatives is our future research.

REFERENCES

Audrino, F., Huitema, R. and Ludwig, M. (2015) An empirical analysis of the Ross recovery theorem. *Available at SSRN: http://ssrn.com/abstract=2433170*.

Backwell, A. (2015) State prices and implementation of the recovery theorem. *Journal of Risk and Financial Management*, **8**, 2-16.

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

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

Dubynskiy, S. and Goldstein, R. S. (2013) Recovering drifts and preference parameters from financial derivatives. *Available at SSRN: http://ssrn.com/abstract= 2244394*.

Martin, I. and Ross, S. A. (2013) The long bond. *Working Paper, Stanford University*.

Park, H. (2015) Ross recovery with recurrent and transient processes. *arXiv preprint arXiv:1410.2282v4*.

Qin, L. and Linetsky, V. (2015) Positive eigenfunctions of Markovian pricing operators: Hansen-Scheinkman factorization and Ross recovery. *arXiv preprint arXiv:1411.3075v2*.

Ross, S. (2015) The recovery theorem. *The Journal of Finance*, **70**, 615-648.

Spears, T. (2013) On estimating the risk-neutral and real-world probability measures. *Doctoral dissertation, Oxford University.*

Walden, J. (2014) Recovery with unbounded diffusion processes. *Available at SSRN: http://ssrn.com/abstract=2508414*.