Estimating Forward Looking Distribution with the Ross Recovery Theorem

Takuya Kiriu*† Norio Hibiki‡

October 27, 2016

Abstract

Recently, Ross [2015] introduced a remarkable theorem, named the "Recovery Theorem." It enables us to estimate the real world distribution from the risk neutral distribution derived from option prices under a particular assumption about a representative investor's risk preferences. The real world distribution estimated with the Recovery Theorem is suitable for many financial problems such as market risk management and portfolio optimization due to its forward looking nature. However, it is not easy to derive the appropriate estimators because of an ill-posed problem in the estimation process. We propose a new method to derive the accurate solution by formulating the regularization term involving prior information. Previous studies propose methods to estimate the real world distribution, but they do not investigate the estimation accuracy. We conduct the numerical analysis with hypothetical data to show the effectiveness of the proposed method. We find the following three points from the results.

1. Stabilizing the solution by introducing the regularization term is an effective method of accurately estimating a real world distribution from the Recovery Theorem.
2. The proposed method can estimate a real world distribution more accurately than the Tikhonov method used by Audrino et al. [2015].
3. Our criteria for selecting a regularization parameter can offer the appropriate parameter in most cases.

* Mitsubishi UFJ Trust Investment Technology Institute Co., Ltd. (MTEC), 4-2-6 Akasaka, Minato-Ku, Tokyo 107-0052, Japan. Corresponding author. E-mail: kiriu@mtec-institute.co.jp

† This research is done in the Graduate School of Science and Technology, Keio University. The views expressed in this paper are those of the authors and do not necessarily reflect the official views of MTEC.

‡ Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku, Yokohama, 223-8522, Japan. E-mail: hibiki@ae.keio.ac.jp

1 Introduction

In general, we need to estimate asset distributions to solve financial problems such as market risk management and optimal asset allocation. A common approach is to estimate the distribution from historical data. However, the financial market is quite volatile, and utilizing a forward looking distribution in implied option prices is more desirable and suitable than a backward looking distribution derived using historical data.

The payoff of an option is determined by the future price of the underlying asset and therefore the option prices contain forward looking information. The forward looking risk neutral distribution can be derived from option prices under the assumption of a complete market (Breeden and Litzenberger [1978]). The risk neutral distribution generally differs from the real world distribution which expresses market participants' consensus. In particular, the expected return (mean) of the risk neutral distribution must be equal to the risk free rate and the expected return of the real world distribution does not have to be. Previous studies proposed methods to adjust a risk neutral distribution to a real world distribution (risk adjustment methods). Bliss and Panigirtzoglou [2004] adjust the distribution assuming CARA utility or CRRA utility as a representative investor's preference. Fackler and King [1990] proposed the method that uses a beta distribution as a calibration function. Shackleton et al. [2010] proposed the nonparametric method that uses kernel density estimation as a calibration function. However, these methods do not offer a completely forward looking real world distribution because the adjustment parameters are estimated from historical data[*1].

Recently, Ross [2015][*2] introduced a remarkable theorem, named the "Recovery Theorem." It enables us to estimate a completely forward looking real world distribution from option prices under a particular assumption about a representative investor's risk preferences. Our paper discusses the method of estimating the distribution using the Recovery Theorem. Ross [2015] shows quite a simple estimation procedure. However, many improvements are required for practical use. Spears [2013], Audrino et al. [2015], Backwell [2015], and Jensen et al. [2015] have developed the practical methodology for estimating the real world distribution from option prices using the Recovery Theorem[*3]. Spears [2013] indicates that estimators derived by the simple 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 the estimation process, and propose application of the Tikhonov method, which is a standard regularization method for ill-posed problems. In addition, they estimate a real

[*1] Kiriu and Hibiki [2014] compare the investment performance of the real world distribution risk-adjusted based on historical data and that of the risk neutral distribution in the asset allocation framework. They show that the risk adjustment using historical data deteriorates investment performance because the appropriate parameter is not obtained from historical data.

[*2] The working paper version of this paper was published in 2011.

[*3] There are other studies related to the Recovery Theorem. Carr and Yu [2012], Dubynskiy and Goldstein [2013], Walden [2014], Park [2015], and Qin and Linetsky [2015] studied the theoretical extension into a continuous time or infinite state. These studies mainly focus on the conditions in which the real world probability can be recovered from the risk neutral probability. Martin and Ross [2013] apply the Recovery Theorem to the long bond, which is a zero coupon bond with infinite maturity, and investigate the result. Borovička et al. [2014] and Bakshi et al. [2015] criticize the assumptions of the Recovery Theorem, but we do not discuss this problem.

world distribution from thirteen years of S&P500 option data and investigate the effectiveness of a simple investment strategy based on the moments of the distribution. Backwell [2015] shows that the time-homogeneity of state prices, which is hypothesized when estimating a real world distribution, cannot be realized in the market. The estimation method is also proposed to reduce the bias. Jensen et al. [2015] generalize the Recovery Theorem by removing the assumption of time-homogeneity. Moreover, they estimate a real world distribution from eighteen years of S&P500 option data and verify the predictive power of the moments. However, the uniqueness of the estimated distribution is not guaranteed.

As shown in Audrino et al. [2015], the regularization method is required to derive the appropriate estimators of the real world distribution because there is an ill-posed problem in the estimation process. However, there are no previous studies which use prior information to solve the ill-posed problem and evaluate the accuracy of the estimation method.

Our contribution is summarized in the following two points.

1. We propose a new method to derive a more accurate solution by formulating the regularization term involving the prior information. Our proposed method provides clear interpretation of the relationship between the regularization parameter and the estimators.
2. We conduct numerical analysis on the estimation accuracy with hypothetical data to show the effectiveness of the proposed method. We find the following three points from the results. (1) Stabilizing the solution by introducing a regularization term is an effective method of accurately estimating a real world distribution with the Recovery Theorem. (2) The proposed method can estimate a real world distribution more accurately than the Tikhonov method. (3) Our criteria for selecting a regularization parameter can offer the appropriate parameter in most cases.

This paper proceeds as follows. Section 2 summarizes the Recovery Theorem of Ross [2015]. Section 3 shows the procedure for estimating the real world distribution from option prices by the Recovery Theorem and proposes the new method. In Section 4, we show the results of the numerical analysis and examine the effectiveness of the proposed method. The final section describes our conclusion and future work.

2 Recovery Theorem

In this section, we summarize the Recovery Theorem of Ross [2015]. We assume an arbitrage free and complete market in discrete time with a finite state, one period model. Market states $\theta_i \, (i = 1, \ldots, n)$ are defined by r_i, the underlying stock index returns from time 0. $P := (p_{i,j})$ is an $n \times n$ transition state price matrix. $p_{i,j}$ is a state price from θ_i to θ_j[*4]. We similarly define an $n \times n$ transition risk neutral probability matrix $Q := (q_{i,j})$ and an $n \times n$ transition real world probability matrix $F := (f_{i,j})$. We also describe the notation Q as "risk neutral probability" and F as "real world probability" depending on the context. Matrix P is assumed

[*4] 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.

to be irreducible[*5], and therefore matrix Q and matrix F are also irreducible. In this section, we suppose that matrix P is known because it can be estimated from option prices[*6]. Matrix Q is easily derived from matrix P, since $q_{i,j}$ is expressed as follows:

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

On the other hand, it is difficult to derive matrix F because the state price is simultaneously a function of both a real world probability and market risk preferences. However, Ross [2015] showed that matrix F can be derived from matrix 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

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

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

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

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

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

Pricing kernel is dependent on investor's risk preferences.

Since matrix P is non-negative and irreducible, the Perron-Frobenius Theorem asserts that matrix P has a unique strictly positive eigenvector $\boldsymbol{v}$ associated with the maximum eigenvalue λ. The Recovery Theorem says that

$$\delta = \lambda \tag{5}$$

$$u'(c_i) = v_i^{-1} \quad (i = 1, \ldots, n) \tag{6}$$

hold, where v_i denotes the i-th element of $\boldsymbol{v}$.

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

$$f_{i,j} = \frac{1}{\lambda} \frac{v_j}{v_i} p_{i,j} \quad (i, j = 1, \ldots, n). \tag{7}$$

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

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

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

3 Implementation of the Recovery Theorem

In this section, we describe the process of estimating the real world distribution with the Recovery Theorem. The process is divided into three steps as referenced by Spears [2013] in Figure 1.

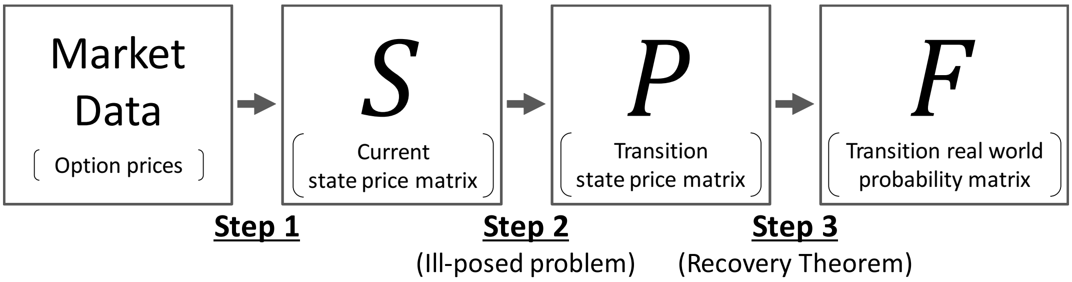


Fig. 1 Estimation steps

An $n \times m$ current state price matrix is defined as $S := (s_{j,\tau})$, where $s_{j,\tau}$ is a current state price for $\tau(= 1, \ldots, m)$ periods transition from current state θ_{i_0} to next state θ_j. Matrix S is estimated from option prices in Step 1. In Step 2, we estimate the $n \times n$ transition state price matrix P from the $n \times m$ matrix S. This section discusses Step 1 and Step 2 because Step 3 simply applies the Recovery Theorem as mentioned in Section 2.

For simplicity, we assume that (1) the number of states n is odd, (2) the current state is the center state ($r_{i_0} = 0$ where $i_0 = (n + 1)/2$), and (3) the return of each state is symmetrical from the center state ($r_{i_0-k} = r_{i_0+k}, (k = 0, \ldots, i_0 - 1)$).

The i_0-th row vector of the matrix P, $\boldsymbol{p}_{i_0}$, represents a state price distribution of the current state. We denote $\boldsymbol{p}_{i_0}$ as the state price distribution. Likewise, we denote $\boldsymbol{q}_{i_0}$ as the risk neutral distribution and $\boldsymbol{f}_{i_0}$ as the real world distribution.

3.1 Step 1: from option prices to matrix S

Breeden and Litzenberger [1978] showed that a state price is calculated from option prices. The state price function $s(k, \tau)$ is represented as follows:

$$s(k, \tau) = \frac{\partial^2 c(k, \tau)}{\partial k^2} \quad (\tau = 1, \ldots, m) \tag{8}$$

where k is the strike price and $c(k, \tau)$ is the call option price function. We can get matrix S by discretizing $s(k, \tau)$. Although this equation is quite simple, there are many methods of estimating the state price function, such as the method of assuming mixed log normal distribution (Melick and Thomas [1997]), the method of using polynomial approximation (Malz [1997]), the method of using a smoothing spline (Bliss and Panigirtzoglou [2002]) and the method of using a neural network (Ludwig [2015]).

However, in our numerical analysis of the estimation accuracy, we generate matrix S from hypothetical data to eliminate the effect of the estimation method of Step 1.

3.2 Step 2: from matrix S to matrix P

In Step 2, we estimate the $n \times n$ matrix P from the $n \times m$ matrix S assuming that state transitions follow a time-homogeneous Markov chain. We assume that $n \leq m$, which means that the number of equations is greater than the number of estimation variables, except for the analysis in Section 4.4.4.

3.2.1 Basic method (Ross [2015])

We explain the basic method of Ross [2015] to estimate matrix P. Denote the first column vector of matrix S by $\boldsymbol{s}_1$. This vector corresponds to the one-period state price distribution of the current state from its definition. It is formulated as,

$$\boldsymbol{s}_1 = \boldsymbol{p}_{i_0}^{\mathsf{T}}. \tag{9}$$

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

$$\boldsymbol{s}_{\tau+1}^{\mathsf{T}} = \boldsymbol{s}_\tau^{\mathsf{T}} P \quad (\tau = 1, \ldots, m-1) \tag{10}$$

Denote the $(m-1) \times n$ matrix transposed from the $n \times m$ matrix S except the last column and the first column respectively by matrix A and B. Equation (10) can be expressed as,

$$AP = B. \tag{11}$$

Matrix P should be estimated by minimizing the differences of both sides of Equation (11) under the no-arbitrage conditions $p_{i,j} \geq 0\,(i, j = 1, \ldots, n)$ and Equation (9). The mathematical formulation is

$$\min_{P} \quad ||AP - B||_2^2 \tag{12}$$

$$\text{subject to} \quad \boldsymbol{s}_1 = \boldsymbol{p}_{i_0}^{\mathsf{T}} \tag{13}$$

$$p_{i,j} \geq 0 \quad (i, j = 1, \ldots, n). \tag{14}$$

3.2.2 Tikhonov method (Audrino et al. [2015])

Audrino et al. [2015] indicate that the average condition number of 11×11 matrix A estimated from S&P 500 option data is very large, and therefore the problem of Equations (12-14) 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 the equations. Consequently, it has the awkward characteristic that the solution is highly sensitive to a small noise. They propose to use the Tikhonov method, which is a standard regularization method for ill-posed problems. The regularization method is formulated by adding the regularization term to the objective function to stabilize the solution. Specifically, the objective function is reformulated as,

$$\min_{P} \quad ||AP - B||_2^2 + \zeta ||P||_2^2 \tag{15}$$

$$\text{subject to} \quad (13) \text{ and } (14).$$

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 (15) can be transformed using an $n \times n$ unit matrix I and an $n \times n$ null matrix O.

$$\min_{P} \quad \left\| \begin{bmatrix} A \\ \sqrt{\zeta} I \end{bmatrix} P - \begin{bmatrix} B \\ O \end{bmatrix} \right\|_2^2 \tag{16}$$

Because the coefficient of matrix P determines the stability of the solution, we focus on the condition number of the matrix created by combining matrices A and $\sqrt{\zeta} I$ vertically. Figure 2 shows the condition number for the hypothetical data explained in Section 4 (base case, $\sigma = 0\%$). The condition number of the original problem (Equation (12) which corresponds to $\zeta = 0$ in the Tikhonov method) is very large, at 1.3×10^{17}. The condition number decreases as ζ increases.

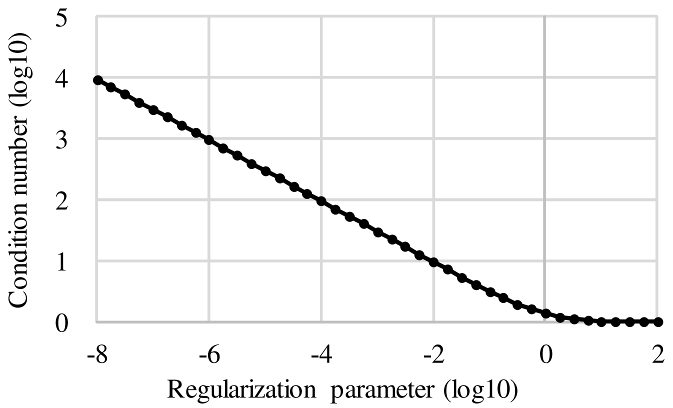


Fig. 2 Condition number with respect to regularization parameter

In the regularization method, we stabilize the solution by adding a regularization term to the objective function. In other words, we derive the solution of the original problem under certain prior information, which is ancillary information about the solution, to solve the problem. It is expected that a solution is derived accurately under appropriate prior information. However, it seems the Tikhonov method does not give the appropriate prior information of this recovery problem. We describe this point as follows. The Tikhonov method stabilizes the solution by the prior information that matrix P is closed to the null matrix as shown in the regularization term of Equation (15). If we set ζ as infinity, $P = O$ is derived. In this case, we cannot get the real world distribution because the Recovery Theorem can apply only to the irreducible matrix[*7]. Therefore, the solution of Step 2 (matrix P) is likely to become stable as ζ increases, whereas the solution of Step 3 (matrix F) is likely to become unstable. According to these properties, the prior information of the Tikhonov method is poorly related with the Recovery Theorem.

Audrino et al. [2015] also proposed the selection method of optimal ζ, which minimizes the discrepancy between original state price matrix (S^O) and the state price matrix implied in

[*7] To be precise, all of the components of the matrix except the i_0-th row become zero because of Constraint (13). Even in this case, matrix P is not irreducible.

matrix P (S^P). The τ-th column vector of S^P is equal to the transposed vector of the i_0-th row of P^τ which is P to the τ-th power($\tau = 1, \ldots, m$). They use the generalized Kullback-Leibler (KL) divergence as a measure of the discrepancy between two matrices. They propose the selection criteria which minimizes the following function $h_A(\zeta)$ defined as

$$h_A(\zeta) := \sum_{i=1}^{n} \sum_{\tau=1}^{m} s^O_{i,\tau} \ln \left(\frac{s^O_{i,\tau}}{s^P_{i,\tau}} \right) - \sum_{i=1}^{n} \sum_{\tau=1}^{m} s^O_{i,\tau} + \sum_{i=1}^{n} \sum_{\tau=1}^{m} s^P_{i,\tau}. \tag{17}$$

Optimal ζ is derived by iterative calculation. However, they do not evaluate the estimation accuracy. Therefore the effectiveness of this selection criteria function has not been examined. We examine the effectiveness of this selection criteria in Section 4.

3.2.3 Proposed method

We propose a new method which modifies the regularization term of the Tikhonov method considering the characteristics of this recovery problem and which has a clear interpretation of the relation between regularization parameter ζ and final estimated value, $\boldsymbol{f}_{i_0}$. We assume the following two different kinds of prior information, PI 1 and PI 2.

PI 1. When we apply the Recovery Theorem to the matrix P where $\zeta = \infty$, the risk neutral distribution is derived as an estimator of the real world distribution.

The conditions that the solution is derived under this prior information are that the matrix P derived for $\zeta = \infty$ is irreducible and the sum of the row elements of the matrix is the same for each row.

PI 2. The risk neutral distribution of the current state, $\boldsymbol{q}_{i_0}$, is close to those of the other states, $\boldsymbol{q}_i (i = 1, \ldots, i_0 - 1, i_0 + 1, \ldots, n)$.

$\boldsymbol{q}_{i_0}$ is determined because of Constraint (13)[*8], while $\boldsymbol{q}_i (i = 1, \ldots, i_0 - 1, i_0 + 1, \ldots, n)$ are unknown. Because it is difficult to set the appropriate values as prior information, we simply assume that $\boldsymbol{q}_{i_0}$ is close to $\boldsymbol{q}_i$, and the risk neutral probabilities whose difference of return between neighboring states is the same are close to each other as [*9],

$$q_{i_0,j} \simeq q_{i_0+k,j+k} \quad (i, j = 1, \ldots, n;\ k \in \mathbb{Z}, 1 \leq i + k \leq n, 1 \leq j + k \leq n). \tag{18}$$

We modify the regularization term based on the two kinds of prior information as follows,

$$\min_P \quad ||AP - B||_2^2 + \zeta ||P - \bar{P}||_2^2 \tag{19}$$

$$\Leftrightarrow \min_P \quad \left\| \begin{bmatrix} A \\ \sqrt{\zeta} I \end{bmatrix} P - \begin{bmatrix} B \\ \sqrt{\zeta} \bar{P} \end{bmatrix} \right\|_2^2 \tag{20}$$

[*8] A risk neutral distribution $\boldsymbol{q}_{i_0}$ is easily calculated from a state price distribution $\boldsymbol{p}_{i_0}$ by Equation (1).

[*9] For example, we assume there are only three states where the returns are $-5\%, 0\%, +5\%$ for each state. The return of a transition from -5% to 0% ($(1+0)/(1-0.05) - 1 = +5.3\%$) is different from that of a transition from 0% to $+5\%$ ($(1+0.05)/(1+0) - 1 = +5\%$). However, both returns are regarded as $+5\%$ approximately.

where,

$$\bar{P} = \begin{bmatrix} \bar{p}_{1,1} & \bar{p}_{1,2} & \cdots & \bar{p}_{1,i_0-1} & \bar{p}_{1,i_0} & \bar{p}_{1,i_0+1} & \cdots & \bar{p}_{1,n-1} & \bar{p}_{1,n} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \bar{p}_{i_0-1,1} & \bar{p}_{i_0-1,2} & \cdots & \bar{p}_{i_0-1,i_0-1} & \bar{p}_{i_0-1,i_0} & \bar{p}_{i_0-1,i_0+1} & \cdots & \bar{p}_{i_0-1,n-1} & \bar{p}_{i_0-1,n} \\ \bar{p}_{i_0,1} & \bar{p}_{i_0,2} & \cdots & \bar{p}_{i_0,i_0-1} & \bar{p}_{i_0,i_0} & \bar{p}_{i_0,i_0+1} & \cdots & \bar{p}_{i_0,n-1} & \bar{p}_{i_0,n} \\ \bar{p}_{i_0+1,1} & \bar{p}_{i_0+1,2} & \cdots & \bar{p}_{i_0+1,i_0-1} & \bar{p}_{i_0+1,i_0} & \bar{p}_{i_0+1,i_0+1} & \cdots & \bar{p}_{i_0+1,n-1} & \bar{p}_{i_0+1,n} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \bar{p}_{n,1} & \bar{p}_{n,2} & \cdots & \bar{p}_{n,i_0-1} & \bar{p}_{n,i_0} & \bar{p}_{n,i_0+1} & \cdots & \bar{p}_{n,n-1} & \bar{p}_{n,n} \end{bmatrix} \tag{21}$$

$$= \begin{bmatrix} \sum_{k=1}^{i_0} s_{k,1} & s_{i_0+1,1} & \cdots & s_{n-1,1} & s_{n,1} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \sum_{k=1}^{2} s_{k,1} & s_{3,1} & \cdots & s_{i_0,1} & s_{i_0+1,1} & s_{i_0+2,1} & \cdots & s_{n,1} & 0 \\ s_{1,1} & s_{2,1} & \cdots & s_{i_0-1,1} & s_{i_0,1} & s_{i_0+1,1} & \cdots & s_{n-1,1} & s_{n,1} \\ 0 & s_{1,1} & \cdots & s_{i_0-2,1} & s_{i_0-1,1} & s_{i_0,1} & \cdots & s_{n-2,1} & \sum_{k=n-1}^{n} s_{k,1} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & s_{1,1} & s_{2,1} & \cdots & s_{i_0-1,1} & \sum_{k=i_0}^{n} s_{k,1} \end{bmatrix}. \tag{22}$$

The problem is subject to the constraints (13-14). The values are accumulated in the first column and last column of matrix $\bar{P}$, and we set to zero the elements which cannot be determined by Equation (18). The proposed method [*10] attempts to find the solution which is close to $\bar{P}$ from the feasible set. The condition number of the problem is the same as that of the Tikhonov method as shown in Figure 2 because the same coefficient of matrix P is used. Therefore, when we use the same ζ in both methods, the sensitivities of the solution to a small noise are also the same for each other.

The proposed method has a clear interpretation about the relation between regularization parameter ζ and the final estimated value, $\boldsymbol{f}_{i_0}$. PI 1 makes all column sums of matrix $\bar{P}$ equal. As we stabilize the solution of Step 2 by increasing ζ, the estimated real world distribution $\boldsymbol{f}_{i_0}$ becomes close to the risk neutral distribution $\boldsymbol{q}_{i_0}$. In other words, the proposed method uses the risk neutral distribution as the basis of estimation. This is methodologically reasonable because the Recovery Theorem is used to get the real world distribution from the risk neutral distribution.

To investigate the relation between the prior information and the estimation accuracy in Section 4.4.3, we formulate the estimation method which assumes only PI 1 or only PI 2, respectively[*11].

[*10] Mathematical formulation involving Equation (19) is called generalized Tikhonov regularization. The proposed method is a special case in which the matrix $\bar{P}$ is defined as Equation (22), while we set $\bar{P} = O$ in the ordinary Tikhonov method.

[*11] We can formulate the estimation method which involves the respective regularization parameters for PI 1 and PI 2. However, we omit it since the analysis is very complicated.

When we solve the problem assuming only PI 1, we replace the objective function (12) with the following function.

$$\min_{P,x} \quad ||AP - B||_2^2 + \zeta ||P - \bar{P}||_2^2 \tag{23}$$

where,

$$\bar{P} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,i_0-1} & x_{1,i_0} & x_{1,i_0+1} & \cdots & x_{1,n-1} & x_{1,n} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ x_{i_0-1,1} & x_{i_0-1,2} & \cdots & x_{i_0-1,i_0-1} & x_{i_0-1,i_0} & x_{i_0-1,i_0+1} & \cdots & x_{i_0-1,n-1} & x_{i_0-1,n} \\ s_{1,1} & s_{2,1} & \cdots & s_{i_0-1,1} & s_{i_0,1} & s_{i_0+1,1} & \cdots & s_{n-1,1} & s_{n,1} \\ x_{i_0+1,1} & x_{i_0+1,2} & \cdots & x_{i_0+1,i_0-1} & x_{i_0+1,i_0} & x_{i_0+1,i_0+1} & \cdots & x_{i_0+1,n-1} & x_{i_0+1,n} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,i_0-1} & x_{n,i_0} & x_{n,i_0+1} & \cdots & x_{n,n-1} & x_{n,n} \end{bmatrix} \tag{24}$$

$$\text{subject to} \quad \sum_{k=1}^{n} x_{i,k} = \sum_{k=1}^{n} s_{k,1} \quad (i = 1, \ldots, i_0 - 1, i_0 + 1, \ldots, n). \tag{25}$$

$x_{i,j}$ $(i = 1, \ldots, i_0 - 1, i_0 + 1, \ldots, n;\ j = 1, \ldots, n)$ are intermediate variables. When we solve the problem assuming only PI 2, we set the objective function (23) and replace the matrix $\bar{P}$ with,

$$\bar{P} = \begin{bmatrix} x_1 \sum_{k=1}^{i_0} s_{k,1} & x_1 s_{i_0+1,1} & \cdots & x_1 s_{n-1,1} & x_1 s_{n,1} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ x_{i_0-1} \sum_{k=1}^{2} s_{k,1} & x_{i_0-1} s_{3,1} & \cdots & x_{i_0-1} s_{i_0,1} & x_{i_0-1} s_{i_0+1,1} & x_{i_0-1} s_{i_0+2,1} & \cdots & x_{i_0-1} s_{n,1} & 0 \\ s_{1,1} & s_{2,1} & \cdots & s_{i_0-1,1} & s_{i_0,1} & s_{i_0+1,1} & \cdots & s_{n-1,1} & s_{n,1} \\ 0 & x_{i_0+1} s_{1,1} & \cdots & x_{i_0+1} s_{i_0-2,1} & x_{i_0+1} s_{i_0-1,1} & x_{i_0+1} s_{i_0,1} & \cdots & x_{i_0+1} s_{n-2,1} & x_{i_0+1} \sum_{k=n-1}^{n} s_{k,1} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & x_n s_{1,1} & x_n s_{2,1} & \cdots & x_n s_{i_0-1,1} & x_n \sum_{k=i_0}^{n} s_{k,1} \end{bmatrix}. \tag{26}$$

x_i $(i = 1, \ldots, i_0 - 1, i_0 + 1, \ldots, n)$ are intermediate variables. Note that in the case of assuming only PI 1 or only PI 2, the sensitivity of the solution of the Tikhonov method is not the same as that of the proposed method even if we use the same value of ζ because the matrix $\bar{P}$ contains variables. Table 1 summarizes the prior information and corresponding formulations with the objective function (23).

The selection method of ζ is also important in finding the accurate solution. The objective function of the optimization problem in Step 2 is Equation (19), and it consists of two terms. The first term shows the fitting error. We denote it by y_{fit}. The second term except ζ, $(||P - \bar{P}||_2^2)$, shows the deviation between the matrices P and $\bar{P}$. We denote it by y_{reg}. Table 3 shows y_{fit} and y_{reg} for the hypothetical data which we explain in Section 4 (base case, $\sigma = 1\%$) for various values of ζ.

As ζ increases, y_{fit} increases and y_{reg} decreases monotonically. Both y_{fit} and y_{reg} have the domain in which the values greatly change. For example, the value of y_{fit} greatly increases around $\log_{10} \zeta = 0$, and y_{reg} decreases around $\log_{10} \zeta = -6$. This is one of the characteristics

Table 1 Prior information and corresponding formulations

Prior Information	ζ	$\bar{P}$	Constraints
None (= Basic method)	$=0$	-	Eqs. (13), (14)
P is close to null matrix (= Tikhonov method)	>0	Null matrix	Eqs. (13), (14)
PI 1	>0	Eq. (24)	Eqs. (13), (14), (25)
PI 2	>0	Eq. (26)	Eqs. (13), (14)
PI 1 and PI 2 (= Proposed method)	>0	Eq. (22)	Eqs. (13),(14)

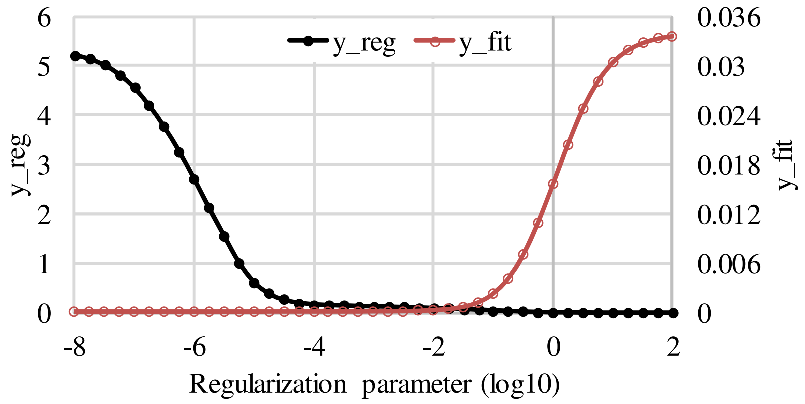


Fig. 3 Decomposition of the objective function: y_{fit} and y_{reg} with respect to the regularization parameter

of the ill-posed problem. The purpose of adding the regularization term to the objective function is to find the optimal solution stably based on the prior information rather than a degenerate solution. We propose a method of selecting ζ by minimizing a selection criteria function $h_K(\zeta)$ defined as,

$$\min_{\zeta} \quad h_K(\zeta) := \frac{y_{fit}(\zeta) - y_{fit}(0)}{y_{fit}(\infty) - y_{fit}(0)} + \frac{y_{reg}(\zeta) - y_{reg}(\infty)}{y_{reg}(0) - y_{reg}(\infty)}. \tag{27}$$

$y_{fit}(\zeta)$ and $y_{reg}(\zeta)$ are functions of ζ as shown in Figure 3. $h_K(\zeta)$ is the sum of the normalized values of y_{fit} and y_{reg}. $y(0)$ is the value without the regularization term and $y(\infty)$ is the value derived under the condition $P = \bar{P}$. Therefore, $y_{reg}(\infty) = 0$ must hold. In addition, $h(0) = 1$ and $h(\infty) = 1$ must hold because both $y_{fit}(\zeta)$ and $y_{reg}(\zeta)$ are monotonic functions. We obtain different values of $h(\zeta)$ by solving the optimization problems for different values of ζ, and then we adopt ζ that minimizes $h_K(\zeta)$.

4 Numerical analysis of estimation accuracy

4.1 Overview

We analyze the estimation accuracy to examine the effectiveness of the proposed method by comparing a preset hypothetical distribution and a distribution which is estimated from the data with noise. We use hypothetical data because it is difficult to specify the true distribution from market data. Using hypothetical data enables us to evaluate estimation accuracy independently of the estimation method of Step 1. Figure 4 shows the overview of the analysis.

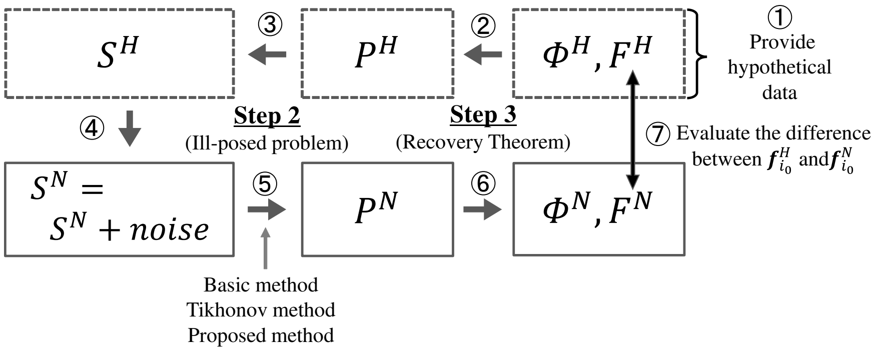


Fig. 4 Overview of analysis

A specific procedure of the analysis is described as follows. Each number corresponds to the number in Figure 4.

① First, we provide two hypothetical matrices: hypothetical real world probability matrix F^H and hypothetical pricing kernel matrix Φ^H.

② A transition state price matrix P^H is calculated backward from the matrices F^H and Φ^H.

③ A current state price matrix S^H is calculated backward from the matrix P^H.

④ We generate a current state price matrix with noise S^N by adding white noise to the matrix S^H. We assume the noise $e_{i,j}$ follows a normal distribution with mean 0 and standard deviation σ. Each component of the matrix S^N is expressed as,

$$s^N_{i,j} = s^H_{i,j}(1 + e_{i,j}) \quad (i, j = 1, \dots, n). \tag{28}$$

⑤ We estimate a matrix P^N from the matrix S^N (Step 2) using the basic method, Tikhonov method, and proposed method.

⑥ A matrix F^N is derived by applying the Recovery Theorem for the matrix P^N.

⑦ If the estimated real world distribution obtained from matrix F^N or $\boldsymbol{f}^N_{i_0}$ is close to $\boldsymbol{f}^H_{i_0}$, the preset real world distribution obtained from the matrix F^H, we evaluate the estimation results as having a high accuracy. A specific evaluation criteria of the estimation accuracy is described in Section 4.2.

4.2 Setting

We explain the setting of the analysis which includes the definition of the state, the number of maturities, evaluation criteria, and comparison of methods.

• State

A market state is defined by the return from time 0. We set 31 returns (states) in total, placed by 2% symmetrically from the return of 0%. Specifically, $r_1 = -30\%$, $r_{16} = 0\%$, and $r_{31} = 30\%$.

- **Number of Maturities**

We can apply any number of maturities m because we calculate the matrix S^H backward from hypothetical data. We set the number of maturities m as equal to n in the base case for simplicity. On the other hand, the number of maturities of options traded in the market is likely to become smaller than the number of states in practice when we estimate a matrix S from market data. We analyze the practical case where $m < n$ in Section 4.4.4[*12].

- **Evaluation criteria**

The estimation accuracy is evaluated by the Kullback-Leibler divergence (KL divergence) of the estimated distribution $\boldsymbol{f}_{i_0}^N$ from the preset hypothetical distribution $\boldsymbol{f}_{i_0}^H$. KL divergence is a measure of the difference between the two distributions and is defined as[*13]

$$D_{KL}(\boldsymbol{f}_{i_0}^H || \boldsymbol{f}_{i_0}^N) := \sum_{j=1}^{n} f_{i_0,j}^N \ln \left(\frac{f_{i_0,j}^N}{f_{i_0,j}^H} \right). \quad (29)$$

If $\boldsymbol{f}_{i_0}^N$ is exactly equal to $\boldsymbol{f}_{i_0}^H$, $D_{KL}(\boldsymbol{f}_{i_0}^H || \boldsymbol{f}_{i_0}^N) = 0$ holds. We also use Euclidean distance instead of KL divergence as a measure of the estimation accuracy, but we arrive at the same conclusion. Therefore, we show only the result using KL divergence hereafter[*14].

- **Comparison of methods**

We compare the estimation accuracy of five methods: "risk neutral (RN) method," "perfect method," and three methods we demonstrated in Section 3.2 (basic method, Tikhonov method, and proposed method).

The RN method uses the risk neutral distribution $\boldsymbol{q}_{i_0}^N$ as an approximation of the real world distribution $\boldsymbol{f}_{i_0}^N$. Risk adjustment by the Recovery Theorem affects the estimation accuracy both positively and negatively. The positive effect is that the risk preference of investors to the distribution can be reflected. The negative effects are that the estimation accuracy is affected by the noise due to the ill-posed problem and it is biased by prior information. From the comparison of the estimation accuracy calculated by each method and the RN method, we evaluate which effect is larger, positive or negative.

The perfect method uses the transition state price matrix P^H calculated from hypothetical data as the prior information $\bar{P}$. In this method, the solution is estimated under the perfectly accurate prior information. Therefore, it is expected that the estimation accuracy monotonically improves as ζ increases. Unfortunately, we cannot use this method practically because we cannot know the true state price matrix P^H when the distribution is estimated from market data. We add this method as a benchmark for comparison to evaluate the estimation accuracy.

[*12] We omit the result of the case where $m > n$, because the conclusion is almost the same as the case where $m = n$.

[*13] We add a very small value (10^{-20}) to each component of $\boldsymbol{f}_{i_0}^H$ and $\boldsymbol{f}_{i_0}^N$ to prevent its anti-logarithm from being zero and avoid dividing it by zero. However, this procedure has no impact on the result.

[*14] We note that the estimation accuracy of the state price distribution $\boldsymbol{s}_{i_0}$ is the same among all methods because of the constraint (13).

4.3 Hypothetical data

The hypothetical data Φ^H and F^H should be generated as appropriately as possible to reproduce the ill-posed problem. We explain the setting of the hypothetical data.

- **Pricing kernel matrix**

We assume that the TAIEUT investor has a CRRA utility function $U(c) = c^{1-\gamma}/(1-\gamma)$ with a relative risk aversion γ. Pricing kernel ϕ is decomposed into U' and δ as shown in Equation (4). We denote the (i,j) element of matrix Φ^H by

$$\phi^H_{i,j} = \delta \left(\frac{1+r_j}{1+r_i} \right)^{-\gamma} \quad (i,j = 1,\ldots,n). \tag{30}$$

The parameters of $\gamma = 3$ and $\delta = 0.999$ are used in the base case*15.

- **Real world probability matrix**

The real world probability matrix F^H is generated based on the S&P500 historical daily price data. We set a reference date and calculate twelve returns in the periods from the reference date to the dates which come every 30 calendar days. If it is a holiday, the return until the day before a holiday is calculated. A matrix is generated by counting the number of state transitions of the return sequence in one period. We denote the return of state θ_j by r_j in the matrix, which is discretely set every 2%. When a real historical return is between $r_j - 1\%$ and $r_j + 1\%$, it is assigned to state θ_j. For example, suppose that a return is 12.5%. It is between $11\%(12\% - 1\%)$ and $13\%(12\% + 1\%)$, and therefore 12% is assigned to the return. A return greater than or equal to 29% (less than or equal to -29%) is 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 the summed matrix is divided by each sum of the row elements to make it a probability matrix. The generated matrix F^H is shown in Table 2.

4.4 Result

4.4.1 Base analysis

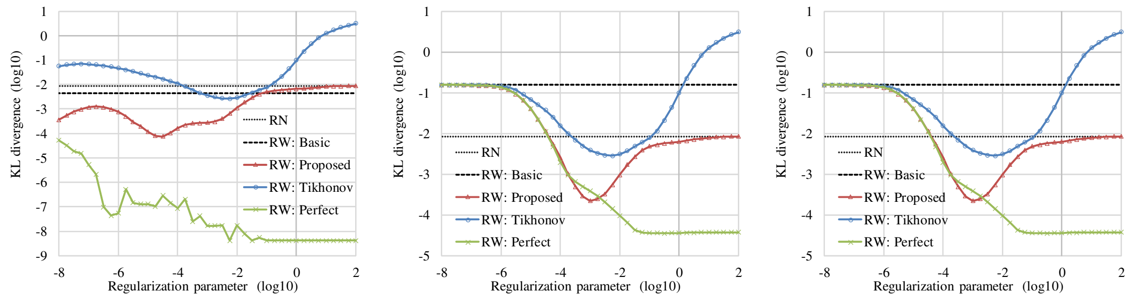
Base analysis compares the estimation accuracy under the setting in Section 4.2 and hypothetical data in Section 4.3. The optimization problem in Step 2 is still ill-posed because the condition number of matrix A^H calculated backward from the matrices Φ^H and F^H is very large, and 1.3×10^{17}. The results for the specific random numbers are shown hereafter, but we obtain the same conclusions for the different random seeds.

Figure 5 displays the KL divergence of $\boldsymbol{f}^N_{i_0}$ from $\boldsymbol{f}^H_{i_0}$ for various values of $\log_{10}\zeta = -8, -7.75, -7.5, \ldots, 2$.

*15 Bliss and Panigirtzoglou [2004] estimate a risk aversion parameter γ implied in S&P500 option data and historical price data from 1993 to 2010. The estimated value is dependent on maturity. The minimum value is 3.37 and the maximum value is 9.52. Therefore, we use $\gamma = 3$ in the base case and $\gamma = 10$ in the robustness check of Section 4.4.2.

Table 2 Hypothetical real world probability matrix F^H generated from the S&P500 historical data

$r_i \setminus r_j$	-30%	-28%	-26%	-24%	-22%	-20%	-18%	-16%	-14%	-12%	-10%	-8%	-6%	-4%	-2%	0%	2%	4%	6%	8%	10%	12%	14%	16%	18%	20%	22%	24%	26%	28%	30%
-30%	0.76	0.07	0.06	0.04	0.03	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-28%	0.31	0.07	0.12	0.18	0.13	0.08	0.06	0.02	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-26%	0.24	0.12	0.06	0.13	0.15	0.12	0.07	0.05	0.03	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-24%	0.20	0.08	0.07	0.06	0.12	0.16	0.15	0.08	0.05	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-22%	0.10	0.04	0.06	0.07	0.12	0.20	0.14	0.15	0.08	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-20%	0.05	0.03	0.04	0.07	0.12	0.15	0.17	0.16	0.11	0.06	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-18%	0.03	0.03	0.03	0.05	0.09	0.13	0.15	0.16	0.15	0.11	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-16%	0.01	0.01	0.03	0.04	0.05	0.09	0.13	0.16	0.16	0.17	0.08	0.03	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-14%	0.01	0.01	0.02	0.02	0.03	0.05	0.08	0.13	0.18	0.17	0.15	0.07	0.04	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-12%	0.00	0.00	0.01	0.02	0.02	0.04	0.04	0.08	0.13	0.16	0.20	0.14	0.08	0.03	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-10%	0.00	0.00	0.00	0.00	0.02	0.02	0.03	0.06	0.09	0.14	0.18	0.18	0.14	0.07	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-8%	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.03	0.06	0.10	0.15	0.17	0.20	0.14	0.07	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-6%	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.06	0.10	0.15	0.19	0.19	0.13	0.07	0.03	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-4%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.06	0.09	0.15	0.18	0.20	0.15	0.07	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-2%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.05	0.09	0.15	0.18	0.21	0.16	0.08	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.08	0.14	0.19	0.23	0.16	0.07	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.08	0.14	0.20	0.23	0.17	0.07	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.08	0.14	0.20	0.24	0.16	0.06	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.08	0.14	0.22	0.24	0.15	0.06	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.08	0.15	0.20	0.23	0.14	0.08	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
10%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.08	0.14	0.20	0.22	0.15	0.08	0.03	0.01	0.00	0.00	0.00	0.00	0.00
12%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.03	0.08	0.14	0.18	0.21	0.18	0.09	0.04	0.01	0.00	0.00	0.00	0.00
14%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.03	0.08	0.13	0.17	0.22	0.17	0.10	0.04	0.01	0.00	0.00	0.00
16%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.03	0.07	0.13	0.19	0.21	0.18	0.09	0.04	0.02	0.01	0.00
18%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.08	0.14	0.19	0.20	0.16	0.10	0.04	0.02	0.01
20%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.08	0.14	0.19	0.19	0.17	0.10	0.04	0.02
22%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.05	0.08	0.13	0.18	0.18	0.16	0.11	0.08
24%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.04	0.09	0.13	0.18	0.18	0.15	0.19
26%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.04	0.08	0.13	0.18	0.18	0.35
28%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.03	0.04	0.08	0.11	0.17	0.53
30%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.03	0.06	0.86



(a)$\sigma = 0\%$ (b)$\sigma = 1\%$ (c)$\sigma = 5\%$

Fig. 5 Base case : KL divergences with respect to regularization parameters

First, we discuss the result of the case of $\sigma = 0\%$ where the matrix S^N is generated without noise. Theoretically, the KL divergence of the basic method is equal to zero. However, the calculated KL divergence of the basic method is 4.4×10^{-3} due to the estimation error. This shows how it is difficult to get an accurate estimator of the ill-posed problem. The estimation accuracy of the proposed method is better than that of the basic method in the range where $\log_{10} \zeta$ is less than 1.5. This result shows that the proposed method is effective to increase the estimation accuracy even without noise. As ζ gets larger, the graph of the proposed method approaches that of the RN method. This is because the real world distribution estimated with the proposed method where $\zeta = \infty$ equals the risk neutral distribution. The estimation accuracy of the Tikhonov method is worse than that of the basic method. The regularization term of the Tikhonov method introduces the bias because it is not formulated with the prior information appropriately. The estimation accuracy of the perfect method is the highest, as we expected. The graph of the perfect method between $\log_{10} \zeta = -6$ and $\log_{10} \zeta = -2$ is distorted by numerical error because the objective value of the optimization problem is very small.

We check the two cases with noise ($\sigma = 1\%$ and $\sigma = 5\%$). The estimation accuracy of the basic method is worse than that of the RN method because of the ill-posed problem. The KL divergence of the Tikhonov method and that of the proposed method are U-shaped in the graph. This indicates that introducing the regularization term is effective to estimate the real world distribution accurately, but the value of ζ needs to be selected appropriately. The estimation accuracy of the proposed method is better than that of the Tikhonov method because the regularization term of the proposed method is formulated involving the prior information more appropriately.

We next examine the effectiveness of the selection function of ζ. Table 3 shows the common logarithm of KL divergence ($\log_{10} D_{KL}$) for each selection criteria of ζ.

Table 3 Base case : Common logarithm of KL divergences ($\log_{10} D_{KL}$) for each selection criteria of regularization parameters

Volatility of noise	Selection criteria of ζ	RN	RW Basic	RW Proposed		RW Tikhonov	
$\sigma = 1\%$	min $h_K(\zeta)$	-2.07	-0.81	-3.18	(-2.23)	-2.07	(-0.89)
	min $h_A(\zeta)$			-1.72	(-4.65)	-0.91	(-5.55)
	min KL			-3.65	(-2.97)	-2.55	(-2.23)
$\sigma = 5\%$	min $h_K(\zeta)$	-2.11	-0.76	-2.57	(-2.39)	-2.03	(-0.97)
	min $h_A(\zeta)$			-1.92	(-2.97)	-0.97	(-3.98)
	min KL			-2.68	(-1.98)	-2.33	(-2.05)

$*$ The common logarithm of $\zeta(\log_{10} \zeta)$ selected by each criteria is in the parenthesis.

The value of "min $h_K(\zeta)$" is calculated by our selection criteria of ζ (Equation (27)) and the value of "min $h_A(\zeta)$" is calculated by the selection criteria proposed by Audrino et al. [2015] (Equation(17)). The values of "min KL" are derived minimizing KL divergence[*16]. We find the optimal value of ζ using the golden section search method in the range from $\log_{10} \zeta = -8$ to $\log_{10} \zeta = 2$. The estimation accuracies of the real world distribution estimated using our proposed method and the Tikhonov method are dependent on the effectiveness of the regularization term of the prior information and the selection criteria of ζ. We can evaluate the effectiveness of the regularization term by comparing the KL divergence of "min KL" of the methods introducing the regulation term with that of the RN method because the value of "min KL" is not dependent on the selection criteria of ζ. As shown in Table 3, the KL divergence of "min KL" is smaller than that of the RN method, and we find that the regulation term can be effective if we select the appropriate criteria of ζ[*17]. Next, we evaluate the selection criteria.

The selection function $h_K(\zeta)$ gives more appropriate ζ than $h_A(\zeta)$ in both the proposed method and the Tikhonov method. The result indicates that it is effective to select ζ based on the normalized value of the residual term and the normalized value of the regularization term. We can derive a more accurate real world distribution by only utilizing the combination

[*16] The parameter ζ for the value of "min KL" is the best because the KL divergence is evaluated as a selection criteria.

[*17] If the KL divergence of "min KL" is larger than that of the RN method, we need to formulate the appropriate regulation term regardless of the selection criteria.

of the proposed method and selection function $h_K(\zeta)$ rather than the RN method.

4.4.2 Robustness check

In this section, we check the robustness of the following three findings obtained in the base analysis: (1) the proposed method and the Tikhonov method are effective to improve the estimation accuracy, (2) we can obtain a more accurate solution by the proposed method than the Tikhonov method, and (3) we can select an appropriate regularization parameter by minimizing the function $h_K(\zeta)$ in the proposed method. The proposed method can be used to derive the real world distribution under two sets of prior information (PI 1 and PI 2). Therefore, it is expected that the estimation accuracy decreases in the case in which the hypothetical data does not reflect the prior information used in the proposed method. We check the robustness using such hypothetical data.

We explain how we provide the hypothetical data which does not reflect the prior information of the proposed method. First, PI 1 is the information that the real world distribution becomes equal to the risk neutral distribution when $\zeta = \infty$. In other word, PI 1 assumes the representative investor is risk neutral ($\gamma = 0$) . As risk aversion increases, the difference between the risk neutral distribution and the real world distribution gets larger. We set a larger risk aversion parameter γ as 10 to check robustness.

Next, PI 2 is the information that the risk neutral distribution of the current state is close to the risk neutral distributions of the other states. A typical example that is different from this information is to set a local volatility, which is the case in which the volatilities of the distributions are largely different in each state. We set the hypothetical data assuming that the real world distributions of each state $\boldsymbol{f}_i (i = 1, \ldots, n)$ follow discretized log normal distributions which have different volatilities $LN(\mu_i, \sigma_i^2)(i = 1, \ldots, n)$. In this case, if the real world probability has a local volatility then the risk neutral probability also has a local volatility. It is a well-known stylized fact that volatilities calculated using the Black Scholes formula have different values for each strike price, which is called volatility smile. Therefore, we can generate the hypothetical data with a local volatility by setting the parameter $\sigma_i (i = 1, \ldots, n)$ based on volatilities implied in market data. We set each mean parameter $\mu_i (i = 1, \ldots, n)$ assuming the risk premium is proportional to volatility. We explain the method specifically.

To check robustness with a high local volatility setting, we generate the hypothetical data based on the implied volatilities on the date when Skew Index (SI), which is the index of the skewness of the S&P500 risk neutral distribution calculated by the Chicago Board Options Exchange (CBOE), takes the highest value. We make the list of dates 30 days before S&P500 option maturities (the third Friday in every month) from Jan 4, 2000 to Jan 12, 2004. Then, we select the date when the highest SI value is shown in the list. The selected date is Jan 22, 2014 and the SI value is 138.92. We calculate implied volatilities from the OTM option with 30-day maturity on Jan 22, 2014. We get an implied volatility $\sigma_i (i = 1, \ldots, n)$ of each state by interpolating with a cubic spline. The implied volatilities are extrapolated with the largest available value because option prices with more than +10% return are not available. Figure 6 shows the implied volatilities σ_i which is related with the return r_i. The volatility is set at a range that is at most five times.

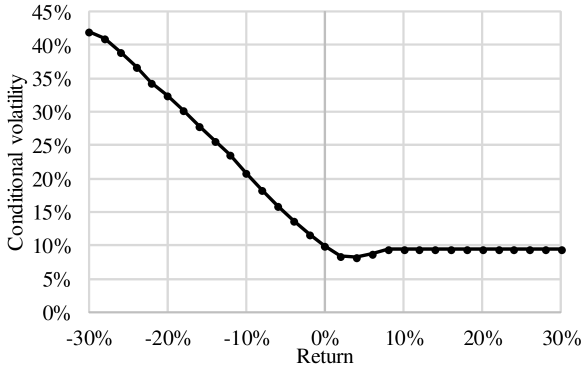


Fig. 6 Implied volatility σ_i calculated from OTM option on Jan 22, 2014.

We assume the risk premium $(\mu_i - r_f)$ is proportional to volatility σ_i as

$$\mu_i - r_f = k\sigma_i \quad (i = 1, \ldots, n), \tag{31}$$

where r_f is a risk free rate and k is a proportionality coefficient. We calculate k

$$k = \frac{\bar{\mu} - \bar{r}_f}{\bar{\sigma}} = 0.02087 \tag{32}$$

where $\bar{\mu}$, $\bar{r}_f$, and $\bar{\sigma}$ are the average of monthly returns of the S&P500, the average of one-month LIBOR, and the average of the CBOE Volatility Index (VIX), respectively. The average is calculated using the monthly data from Jan, 1990 to Dec, 2014. Next, we calculate μ_i from Equation (31) where r_f is the one-month LIBOR on Jan 22, 2014. A matrix F^H is generated by discretizing the log-normal distribution and it is shown in Table 4.

Table 4 Hypothetical real world probability matrix F^H with a local volatility

$r_i \setminus r_j$	-30%	-28%	-26%	-24%	-22%	-20%	-18%	-16%	-14%	-12%	-10%	-8%	-6%	-4%	-2%	0%	2%	4%	6%	8%	10%	12%	14%	16%	18%	20%	22%	24%	26%	28%	30%
-30%	0.52	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-28%	0.42	0.09	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-26%	0.33	0.09	0.10	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-24%	0.24	0.09	0.10	0.10	0.09	0.09	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-22%	0.15	0.08	0.09	0.10	0.10	0.10	0.09	0.08	0.06	0.05	0.04	0.02	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-20%	0.09	0.06	0.08	0.09	0.10	0.11	0.10	0.09	0.08	0.06	0.05	0.03	0.02	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-18%	0.04	0.04	0.06	0.08	0.09	0.11	0.11	0.11	0.09	0.08	0.06	0.05	0.03	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-16%	0.02	0.02	0.03	0.05	0.08	0.10	0.11	0.12	0.11	0.10	0.08	0.06	0.04	0.03	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-14%	0.00	0.01	0.02	0.03	0.05	0.08	0.10	0.12	0.12	0.12	0.10	0.08	0.06	0.04	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-12%	0.00	0.00	0.01	0.01	0.03	0.05	0.08	0.10	0.12	0.13	0.13	0.11	0.08	0.06	0.04	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-10%	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.08	0.11	0.14	0.15	0.14	0.11	0.08	0.06	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-8%	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.07	0.11	0.15	0.16	0.15	0.12	0.08	0.05	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-6%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.06	0.11	0.16	0.18	0.17	0.13	0.08	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-4%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.05	0.11	0.18	0.21	0.18	0.13	0.07	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-2%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.11	0.19	0.24	0.20	0.13	0.06	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.10	0.21	0.27	0.22	0.12	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.08	0.22	0.31	0.24	0.10	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.08	0.22	0.31	0.24	0.10	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.09	0.21	0.29	0.23	0.11	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.10	0.20	0.26	0.22	0.12	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.10	0.20	0.26	0.22	0.12	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00
12%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.10	0.20	0.25	0.21	0.12	0.05	0.01	0.00	0.00	0.00	0.00	0.00
14%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.10	0.20	0.25	0.21	0.12	0.05	0.01	0.00	0.00	0.00	0.00
16%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.10	0.20	0.25	0.21	0.12	0.05	0.02	0.00	0.00	0.00
18%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.11	0.19	0.24	0.21	0.13	0.05	0.02	0.00	0.00
20%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.11	0.19	0.24	0.21	0.13	0.06	0.02	0.01
22%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.11	0.19	0.23	0.20	0.13	0.06	0.03
24%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.11	0.19	0.23	0.20	0.13	0.09
26%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.11	0.19	0.23	0.20	0.22
28%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.05	0.11	0.19	0.22	0.42
30%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.05	0.11	0.18	0.64

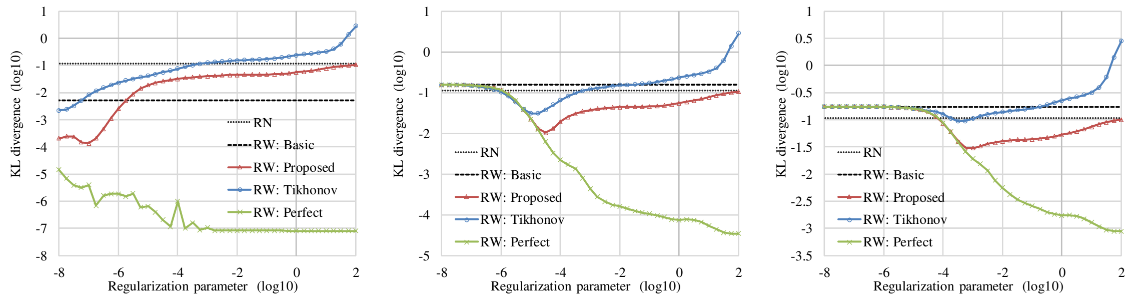
Next, we conduct the analysis for the three new cases shown in Table 5. We show the results for the three cases.

Table 5　Robustness check : Setting of the hypothetical data

		Real world matrix F^H	
		Table 2	Table 4
Pricing kernel matrix Φ^H	$\gamma = 3$	Base case	Case B
	$\gamma = 10$	Case A	Case C

- **Case A**

In case A, the pricing kernel matrix Φ^H where $\gamma = 10$ and the real world probability matrix F^H generated based on historical data are used as the hypothetical data for the analysis. Figure 7 shows the KL divergences with respect to ζ.



(a)$\sigma = 0\%$　(b)$\sigma = 1\%$　(c)$\sigma = 5\%$

Fig. 7　Case A : KL divergences with respect to regularization parameters

The main characteristics are the same as in the base case. The proposed method and the Tikhonov method improve the estimation accuracy by stabilizing the solution of the ill-posed problem. The KL divergence of the proposed method is less than that of the Tikhonov method regardless of ζ level.

Table 6 shows the KL divergences for the three types of selection criteria of ζ.

Table 6　Case A : Common logarithm of KL divergences ($\log_{10} D_{KL}$) for each selection criteria of regularization parameters

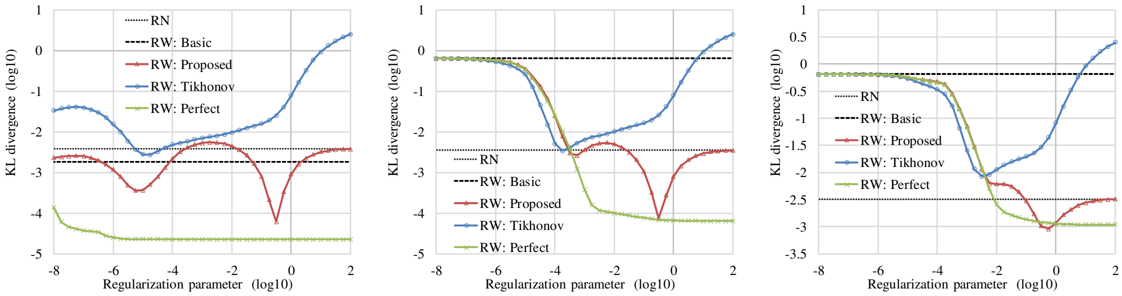
Volatility of noise	Selection criteria of ζ	RN	RW Basic	RW Proposed		RW Tikhonov	
$\sigma = 1\%$	min $h_K(\zeta)$	-0.95	-0.81	-1.35	(-1.46)	-0.65	(-0.14)
	min $h_A(\zeta)$			-1.55	(-5.06)	-0.81	(-7.39)
	min KL			-1.98	(-4.47)	-1.52	(-4.88)
$\sigma = 5\%$	min $h_K(\zeta)$	-0.96	-0.76	-1.38	(-1.72)	-0.71	(-0.38)
	min $h_A(\zeta)$			-0.97	(-4.14)	-0.78	(-5.32)
	min KL			-1.53	(-3.06)	-1.04	(-3.39)

$*$ The common logarithm of $\zeta(\log_{10}\zeta)$ selected by each criteria is in the parenthesis.

The estimation accuracy of the combination of the Tikhonov method and the selection function $h_A(\zeta)$ is inferior to that of the RN method. On the other hand, the estimation accuracy of the combination of the proposed method and the selection function $h_K(\zeta)$ is superior to that of the RN method. In case A, the three findings obtained from the base case are observed.

- **Case B**

We utilize the hypothetical real world probability from the matrix generated with a local volatility (Table 4) in case B, instead of the matrix generated based on historical data (Table 2). The pricing kernel matrix is the same as that of the base case. Figure 8 shows the KL divergences with respect to ζ.



(a)$\sigma = 0\%$ (b)$\sigma = 1\%$ (c)$\sigma = 5\%$

Fig. 8 Case B : KL divergences with respect to regularization parameters

The proposed method and the Tikhonov method reduce the KL divergence. The KL divergence of the proposed method sharply drops off around $\log_{10}\zeta = -0.5$. This is because the best solution exists between the solutions of the basic method and the RN method, and the solution of the proposed method is close to the best solution.

Table 7 shows the KL divergences for the three selection criteria of ζ.

Table 7 Case B : Common logarithm of KL divergences ($\log_{10} D_{KL}$) for each selection criteria of regularization parameters

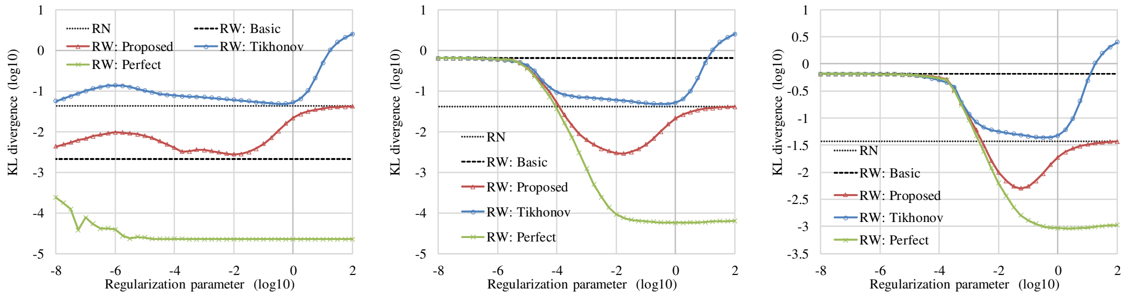
Volatility of noise	Selection criteria of ζ	RN	RW: Basic	RW: Proposed		RW: Tikhonov	
$\sigma = 1\%$	min $h_K(\zeta)$	-2.45	-0.20	-2.50	(-3.06)	-1.78	(-0.97)
	min $h_A(\zeta)$			-0.26	(-5.73)	-0.21	(-6.89)
	min KL			-4.13	(-0.48)	-2.47	(-3.73)
$\sigma = 5\%$	min $h_K(\zeta)$	-2.49	-0.19	-0.93	(-3.19)	-1.72	(-1.15)
	min $h_A(\zeta)$			-0.30	(-4.39)	-0.19	(-8.50)
	min KL			-2.90	(-0.38)	-2.07	(-2.52)

$*$ The common logarithm of $\zeta(\log_{10}\zeta)$ selected by each criteria is in the parenthesis.

The combination of the proposed method and the selection function $h_K(\zeta)$ can estimate a more accurate solution than the others in most cases, but the estimation accuracy of this combination is worse than that of the RN method for $\sigma = 5\%$. However, it is expected that a more accurate solution can be derived by improving the selection function because the minimum KL divergence of the proposed method is lower than the KL divergence of the RN method.

- **Case C**

In case C, we use the matrix Φ^H where $\gamma = 10$ and the matrix F^H showed in Table 4 as hypothetical data. Figure 9 shows the KL divergences with respect to ζ.



(a)$\sigma = 0\%$ (b)$\sigma = 1\%$ (c)$\sigma = 5\%$

Fig. 9 Case C : KL divergences with respect to regularization parameters

It may be anticipated that the estimation accuracy of the proposed method largely decreases because we use the hypothetical data which does not reflect both PI 1 and PI 2. In fact, the result is almost the same as the base case. In the cases with noise, the estimation accuracy of the proposed method largely increases when $\log_{10}\zeta$ is around -2 and is better than the estimation accuracy of the Tikhonov method and RN method.

Table 8 shows the KL divergences for the three selection criteria of ζ.

Table 8 Case C : Common logarithm of KL divergences ($\log_{10} D_{KL}$) for each selection criteria of regularization parameters

Volatility of noise	Selection criteria of ζ	RN	RW: Basic	RW: Proposed		RW: Tikhonov	
$\sigma = 1\%$	min $h_K(\zeta)$	-1.38	-0.19	-2.53	(-1.98)	-1.32	(-0.41)
	min $h_A(\zeta)$			-0.26	(-5.39)	-0.20	(-7.45)
	min KL			-2.54	(-1.82)	-1.32	(-0.38)
$\sigma = 5\%$	min $h_K(\zeta)$	-1.42	-0.19	-1.89	(-2.13)	-1.36	(-0.48)
	min $h_A(\zeta)$			-0.24	(-4.14)	-0.19	(-11.96)
	min KL			-2.30	(-1.25)	-1.36	(-0.48)

$*$ The common logarithm of $\zeta(\log_{10}\zeta)$ selected by each criteria is in the parenthesis.

The combination of the proposed method and the selection function $h_K(\zeta)$ gives a more accurate solution than the RN method and the combination of the Tikhonov method and the selection function $h_A(\zeta)$.

Why can the proposed method give an accurate estimator under the hypothetical data, regardless of the inappropriate prior information of the proposed method? This is because the first term of the objective function (19) weighs more heavily than the second term associated with the prior information if the proposed method cannot use the appropriate prior information to derive the solution stably. Therefore, we can derive the accurate estimator. Even in this case, it is important to involve the appropriate regularization term considering the characteristics of the Recovery Theorem because the estimation accuracy of the proposed method is superior to that of the Tikhonov method.

- **Robustness check for other parameters**

We have checked the robustness of the result in four cases of hypothetical data so far. In addition, we evaluate the sensitivity of the following parameters to check the robustness.

Data: We analyze the sensitivity with respect to the hypothetical real world probability matrix F^H generated based on the Nikkei 225 instead of the S&P500 under the base case and case A. We use daily data from Jan 4, 1950 to Jan 4, 2014.

μ : We assume the constant mean μ instead of the state-dependent mean $\mu_i (i = 1, \ldots, n)$ which is estimated under the matrix F^H with a local volatility. We set μ as the average of $\mu_i (i = 1, \ldots, n)$. We analyze the sensitivity under case B and case C.

$U(c)$: We assume the investor has a CARA utility function ($U(c) = -\exp(-\gamma x)/\gamma$) instead of CRRA. In this case the hypothetical pricing kernel matrix is generated by

$$\phi^H_{i,j} = \delta e^{-\gamma(r_j - r_i)} \quad (i, j = 1, \ldots, n). \tag{33}$$

δ : We change the discount factor from $\delta = 0.999$ to $\delta = 0.99$.

n : We change the number of states from 31 to 61. Specifically, $r_1 = -30\%$, $r_{31} = 0\%$, and $r_{61} = 30\%$.

Tables 9 to 12 show the KL divergences in the cases where only underlined parameters are changed from the base case, case A, case B, and case C, respectively.

The minimum KL divergences of the real world distribution estimated by the proposed method are less than those of the risk neutral distribution in all sixteen cases where $\sigma = 1\%$ and in fourteen of sixteen cases where $\sigma = 5\%$. In addition, the minimum KL divergences of the proposed method are less than those of the Tikhonov method in all 32 cases. The real world distribution estimated by the combination of the proposed method and selection criteria $h_K(\zeta)$ gives more accurate estimators than the risk neutral distribution in fifteen cases where $\sigma = 1\%$ and nine cases where $\sigma = 5\%$. The three findings of the base analysis are observed in most cases.

We cannot fully show the robustness of the results because of the limited hypothetical datasets. However, we obtain almost the same results as the base analysis even if we use the hypothetical data which does not reflect the prior information of the proposed method. It is expected that we will get similar results in other cases. The further analysis is our future work.

Table 9　Sensitivity analysis of base case： Common logarithm of KL divergences

Setting					RN	RW:Basic	RW:Proposed			RW:Tikhonov		
Data	$U(c)$	δ	n	σ			min $h_K(\zeta)$	min $h_A(\zeta)$	min KL	min $h_K(\zeta)$	min $h_A(\zeta)$	min KL
Nikkei225	CRRA	0.999	31	1%	-1.86	-0.78	-1.73	-1.30	-2.38	-1.62	-0.97	-2.09
				5%	-1.88	-0.75	-1.78	-1.75	-1.87	-1.64	-1.09	-1.83
S&P500	CARA	0.999	31	1%	-2.08	-0.81	-3.38	-1.85	-3.79	-1.99	-0.91	-2.56
				5%	-2.12	-0.76	-2.47	-1.83	-2.70	-1.96	-0.95	-2.33
S&P500	CRRA	0.99	31	1%	-2.07	-0.80	-3.20	-1.57	-3.58	-1.94	-0.91	-2.55
				5%	-2.11	-0.76	-2.47	-1.64	-2.67	-1.94	-0.93	-2.32
S&P500	CRRA	0.999	61	1%	-1.84	-0.82	-1.84	-1.52	-2.50	-1.53	-0.97	-2.15
				5%	-1.81	-0.93	-1.75	-1.78	-1.81	-1.53	-1.42	-1.76

∗ Underline represents the parameter which is different from the original base case

Table 10　Sensitivity analysis of case A： Common logarithm of KL divergences

Setting					RN	RW:Basic	RW:Proposed			RW:Tikhonov		
Data	$U(c)$	δ	n	σ			min $h_K(\zeta)$	min $h_A(\zeta)$	min KL	min $h_K(\zeta)$	min $h_A(\zeta)$	min KL
Nikkei225	CRRA	0.999	31	1%	-0.75	-0.79	-1.26	-1.64	-1.66	-0.22	-0.72	-1.02
				5%	-0.76	-0.71	-1.29	-1.13	-1.49	-0.28	-0.78	-0.80
S&P500	CARA	0.999	31	1%	-0.99	-0.80	-1.28	-1.41	-1.87	-0.73	-0.92	-1.52
				5%	-1.00	-0.76	-1.34	-0.81	-1.46	-0.76	-0.77	-1.02
S&P500	CRRA	0.99	31	1%	-0.95	-0.81	-1.35	-1.37	-1.89	-0.68	-1.01	-1.50
				5%	-0.96	-0.76	-1.39	-0.92	-1.51	-0.72	-0.78	-1.01
S&P500	CRRA	0.999	61	1%	-0.74	-0.84	-1.22	-1.28	-1.51	-0.07	-0.99	-1.02
				5%	-0.74	-0.93	-1.25	-1.16	-1.38	-0.16	-0.57	-0.95

∗ Underline represents the parameter which is different from the original case A

Table 11　Sensitivity analysis of case B： Common logarithm of KL divergences

Setting					RN	RW:Basic	RW:Proposed			RW:Tikhonov		
μ	$U(c)$	δ	n	σ			min $h_K(\zeta)$	min $h_A(\zeta)$	min KL	min $h_K(\zeta)$	min $h_A(\zeta)$	min KL
Cons	CRRA	0.999	31	1%	-2.45	-0.19	-2.57	-0.25	-4.21	-1.72	-0.21	-2.18
				5%	-2.49	-0.13	-0.84	-0.32	-3.07	-1.65	-0.14	-1.90
Prop	CARA	0.999	31	1%	-2.45	-0.20	-2.47	-0.32	-4.09	-1.78	-0.22	-2.46
				5%	-2.49	-0.19	-0.95	-0.32	-3.04	-1.71	-0.19	-2.06
Prop	CRRA	0.99	31	1%	-2.45	-0.20	-2.51	-0.31	-4.12	-1.77	-0.20	-2.43
				5%	-2.49	-0.16	-0.95	-0.26	-3.05	-1.72	-0.17	-2.05
Prop	CRRA	0.999	61	1%	-2.43	-0.23	-2.96	-0.49	-4.27	-1.80	-0.34	-2.55
				5%	-2.37	-0.30	-0.99	-0.68	-3.12	-1.74	-0.31	-1.92

∗ Underline represents the parameter which is different from the original case B

† "Cons" represents constant and "Prop" represents proportional

Table 12　Sensitivity analysis of case C： Common logarithm of KL divergences

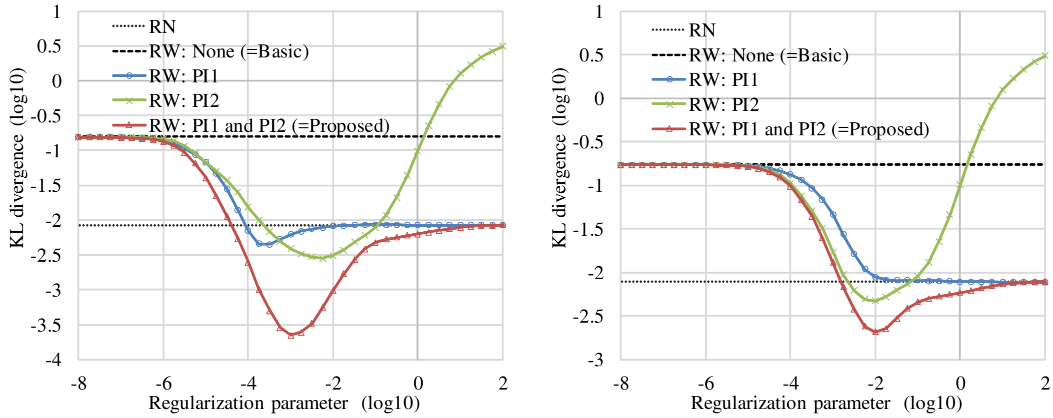
Setting					RN	RW:Basic	RW:Proposed			RW:Tikhonov		
μ	$U(c)$	δ	n	σ			min $h_K(\zeta)$	min $h_A(\zeta)$	min KL	min $h_K(\zeta)$	min $h_A(\zeta)$	min KL
Cons	CRRA	0.999	31	1%	-1.38	-0.19	-2.32	-0.38	-2.35	-1.46	-0.21	-1.46
				5%	-1.42	-0.12	-1.99	-0.20	-2.12	-1.50	-0.12	-1.50
Prop	CARA	0.999	31	1%	-1.38	-0.20	-2.32	-0.62	-2.38	-1.31	-0.21	-1.32
				5%	-1.42	-0.19	-1.46	-0.27	-2.15	-1.35	-0.19	-1.35
Prop	CRRA	0.99	31	1%	-1.38	-0.20	-2.52	-0.23	-2.55	-1.31	-0.21	-1.31
				5%	-1.42	-0.16	-1.89	-0.24	-2.28	-1.35	-0.17	-1.35
Prop	CRRA	0.999	61	1%	-1.39	-0.23	-2.41	-0.47	-2.57	-1.38	-0.22	-1.40
				5%	-1.39	-0.30	-2.02	-0.80	-2.14	-1.36	-0.31	-1.38

∗ Underline represents the parameter which is different from the original case C

† "Cons" represents constant and "Prop" represents proportional

4.4.3 Effect of the prior information

This section analyzes the contribution of the prior information to the estimation accuracy. In particular, we compare the estimation accuracy among five formulations: four methods without the Tikhonov method shown in Table 1 and the RN method. Figure 10 shows KL divergences with respect to ζ in the setting of the base case.



(a)$\sigma = 1\%$ (b)$\sigma = 5\%$

Fig. 10 Effect of the prior information: KL divergences with respect to the regularization parameters

In both the case of $\sigma = 1\%$ and 5%, the graph of “PI 1” approaches the risk neutral distribution. The KL divergence of “PI 1” is less than that of RN method around $\log_{10} \zeta = -3.5$ and the KL divergence of “PI 2” is less than that of RN method around $\log_{10} \zeta = -2$. We must note that the sensitivities of the solutions of the three formulations (“PI 1”, “PI 2”, and “PI 1 and PI 2”) are different because the variables are included in the regularization terms of “PI 1” and “PI 2.” Therefore we evaluate the estimation accuracy by the minimum KL divergence as shown in Table 13.

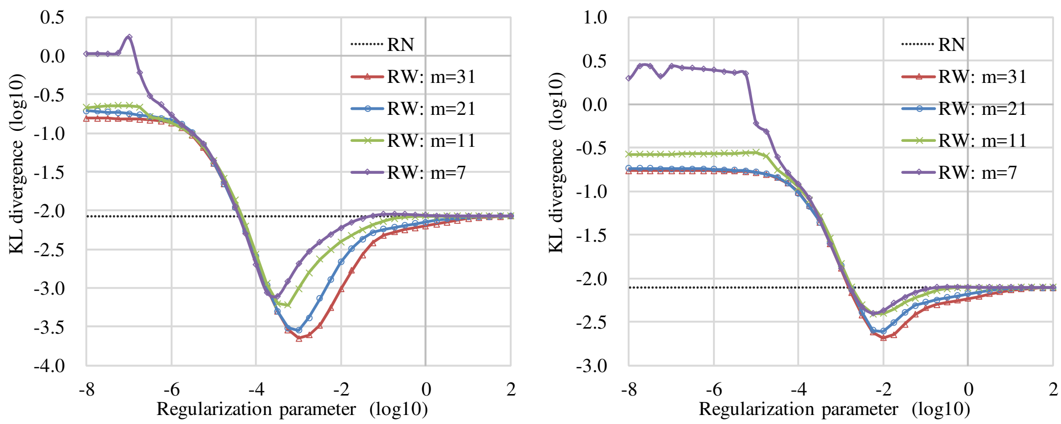
Table 13 Effect of the prior information: Common logarithm of minimum KL divergence ($\log_{10} D_{KL}$)

	$\sigma = 1\%$	$\sigma = 5\%$
RN	-2.07	-2.11
RW: None (= Basic method)	-0.81	-0.76
RW: PI 1	-2.36	-2.10
RW: PI 2	-2.55	-2.33
RW: PI 1 and PI 2 (= Proposed method)	-3.65	-2.68

The formulations can be ranked in the estimation accuracy as “PI 1 and PI 2”, “PI 2”, “PI 1”, “RN”, and “None”. This result shows that both “PI 1” and “PI 2” contribute to the improvement of estimation accuracy and the proposed method can estimate more accurate solutions by the combination of “PI 1” and “PI 2.”

4.4.4 Effect of insufficient data

We have analyzed the estimation accuracy with $m = 31$, where m is the number of option maturities traded in the market. In practice, the number of option maturities is less than 31. For example, the number of S&P500 option maturities traded monthly in the CBOE is twelve. The number of Nikkei 225 options in the Osaka Exchange is nine. In addition, the number m becomes smaller because long-term options are likely to have low liquidity. We conduct an analysis for the case where the number of maturities (equations) m is less than the number of states (estimated variables) n. Specifically, we estimate the real world distribution where $n = 31$ and there are four kinds of the numbers of column of S^H $(m = 7, 11, 21, 31)$ by the proposed method and calculate the KL divergences. Figure 11 shows KL divergences with respect to the regularization parameters.



(a)$\sigma = 1\%$ (b)$\sigma = 5\%$

Fig. 11 Effect of insufficient data : KL divergences with respect to the regularization parameters

Usually, it is difficult to get an accurate estimator because the solution is not determined uniquely when the number of equations m is less than that of estimated parameters n. Figure 11 shows that the estimation accuracy decreases as m decreases. However, the KL divergence of the proposed method where $m = 7$ is less than that of the RN method around $\log_{10} \zeta = -3.5$ for $\sigma = 1\%$ and $\log_{10} \zeta = -2$ for $\sigma = 5\%$. This is because the prior information included in the regularization term offsets the insufficient information. In other words, the necessary information to estimate the real world distribution is almost included in the state price matrix of seven maturities.

In addition, we evaluate the estimation accuracy with respect to the selection of ζ. Table 14 shows the KL divergence for each selection criteria.

In the case of $\sigma = 1\%$, the KL divergence of the proposed method with $h_K(\zeta)$ is less than that of the RN method even if $m = 7$. On the other hand, the KL divergence of $h_A(\zeta)$ increases by the risk adjustment because $h_A(\zeta)$ cannot select an appropriate ζ. In the case of $\sigma = 5\%$, the KL divergence of the proposed method with $h_K(\zeta)$ is not less than that of the RN method except for $m = 31$ which has a sufficient number of equations. However, it is expected that the accurate solution is obtained by setting more appropriate selection criteria because the minimum KL divergence is less than the KL divergence of the RN method.

Table 14 Effect of insuffient data : Common logarithm of KL divergences ($\log_{10} D_{KL}$) for each selection criteria of regularization parameters

Volatility of noise	Selection criteria of ζ	RN	RW			
			$m = 31$	$m = 21$	$m = 11$	$m = 7$
$\sigma = 1\%$	min $h_K(\zeta)$	-2.07	-3.18	-3.32	-3.16	-2.67
	min $h_A(\zeta)$		-1.72	-1.14	-0.79	-0.18
	min KL		-3.65	-3.57	-3.24	-3.13
$\sigma = 5\%$	min $h_K(\zeta)$	-2.11	-2.57	-2.07	-1.23	-0.88
	min $h_A(\zeta)$		-1.92	-1.24	-0.56	0.00
	min KL		-2.68	-2.63	-2.42	-2.40

5 Conclusion

The Recovery Theorem of Ross [2015] enables us to estimate the real world distribution from the risk neutral distribution. However, it is not easy to derive an appropriate estimator because there is an ill-posed problem in the estimation process. We proposed a new method to derive the appropriate estimator by formulating the regularization term involving the prior information. The estimated real world distribution of the proposed method approaches the risk neutral distribution by increasing the regularization parameter, and therefore the solution can be stable. It is important to interpret the relationship between the regularization parameter and the estimator clearly from a practical perspective when we utilize the estimation method.

We compared the estimation accuracy of the proposed method with that of the Tikhonov method in our numerical analysis. From the result, we found the following three points: (1) the method of introducing the regularization term is effective to estimate a real world distribution accurately and stably, (2) the proposed method can estimate a real world distribution more accurately than the Tikhonov method, and (3) the proposed criteria to select the regularization parameter can give the appropriate parameter in most cases. We check the robustness using the hypothetical data which does not reflect the prior information of the proposed method.

Our future work is to estimate the real world distribution from the market option price, and examine the investment performance and the empirical effectiveness of the proposed method by statistical test[*18]. Zdorovenin and Pézier [2011] and Kiriu and Hibiki [2014] derive the real world distribution with risk adjustment which uses historical data and compare the investment performance of the real world distribution with that of the risk neutral distribution. Forward looking risk adjustment using the Recovery Theorem will enhance the investment performance of backward looking risk adjustment using historical data.

[*18] Recently, Flint and Mare [2016] implemented our proposed method based on the working paper version of this study. They estimate the real world distribution with our proposed method from South African stock futures options and examine the investment performance of a simple investment strategy of Audrino et al. [2015]. They show that the investment performance of the real world distribution estimated with our proposed method is superior to that of the risk neutral distribution. However, they do not conduct the statistical test.

References

F. Audrino, R. Huitema, and M. Ludwig. An empirical analysis of the ross recovery theorem. *Working Paper, Available at SSRN 2433170*, 2015.

A. Backwell. State prices and implementation of the recovery theorem. *Journal of Risk and Financial Management*, 8(1):2–16, 2015.

G. Bakshi, F. Chabi-Yo, and X. Gao. A recovery that we can trust? deducing and testing the restrictions of the recovery theorem. *Working Paper*, 2015.

R. R. Bliss and N. Panigirtzoglou. Testing the stability of implied probability density functions. *Journal of Banking & Finance*, 26(2):381–422, 2002.

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

J. Borovička, L. P. Hansen, and J. A. Scheinkman. Misspecified recovery. Technical report, National Bureau of Economic Research, 2014.

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

P. Carr and J. Yu. Risk, return, and ross recovery. *Journal of Derivatives*, 20(1):38, 2012.

S. Dubynskiy and R. S. Goldstein. Recovering drifts and preference parameters from financial derivatives. *Working Paper, Available at SSRN 2244394*, 2013.

P. L. Fackler and R. P. King. Calibration of option-based probability assessments in agricultural commodity markets. *American Journal of Agricultural Economics*, 72(1):73–83, 1990.

E. J. Flint and E. Mare. Estimating option-implied distributions in illiquid markets and implementing the ross recovery theorem. *Working Paper, Available at SSRN 2817080*, 2016.

C. S. Jensen, D. Lando, and L. H. Pedersen. Generalized recovery. *Working Paper, Available at SSRN 2674541*, 2015.

T. Kiriu and N. Hibiki. Optimal asset allocation model with implied distributions for multiple assets. *Transactions of the Operations Research Society of Japan*, 57:112–134, 2014. (In Japanese).

M. Ludwig. Robust estimation of shape-constrained state price density surfaces. *The Journal of Derivatives*, 22(3):56–72, 2015.

A. M. Malz. Estimating the probability distribution of the future exchange rate from option prices. *The Journal of Derivatives*, 5(2):18–36, 1997.

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

W. R. Melick and C. P. Thomas. 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(1):91–115, 1997.

H. Park. Ross recovery with recurrent and transient processes. *Working Paper, arXiv preprint arXiv:1410.2282*, 2015.

L. Qin and V. Linetsky. Positive eigenfunctions of markovian pricing operators: Hansen-scheinkman factorization and ross recovery. *Working Paper, arXiv preprint arXiv:1411.3075*, 2015.

S. Ross. The recovery theorem. *The Journal of Finance*, 70(2):615–648, 2015.

M. B. Shackleton, S. J. Taylor, and P. Yu. A multi-horizon comparison of density forecasts for the s&p 500 using index returns and option prices. *Journal of Banking & Finance*, 34 (11):2678–2693, 2010.

T. Spears. On estimating the risk-neutral and real-world probability measures. *PhD thesis, Oxford University*, 2013.

J. Walden. Recovery with unbounded diffusion processes. *Working Paper, Available at SSRN 2508414*, 2014.

V. V. Zdorovenin and J. Pézier. Does the information content of option prices add value for asset allocation? *ICMA Centre Discussion Paper No. DP2011-03*, 2011.